

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2001, 05:57:59 ; Search time 2308.95 Seconds
(without alignments)
4322.143 Million cell updates/sec

Title: US-09-653-961-1

Perfect score: 1950
Sequence: 1 atggggctccaccagctgtg.....tgaggcattagccccaagt 1950

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_da1:*
2: gb_da2:*
3: gb_cm:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pl2:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_in:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vl:*
29: gb_da3:*
30: gb_in1:*
31: gb_in2:*
32: gb_in3:*
33: gb_pl3:*
34: gb_pr4:*
35: em_da1:*
36: em_da2:*
37: em_hg1:*
38: em_hg2:*
39: em_hg3:*
40: em_hg4:*
41: em_hg5:*
42: em_hg6:*
43: em_hg7:*

44: em_hg8:*
45: em_hg9:*
46: em_hg10:*
47: em_hum3:*
48: em_hum4:*
49: em_hum5:*
50: em_hum6:*
51: gb_pr5:*
52: gb_pr6:*
53: gb_pr7:*
54: gb_hg1:*
55: gb_hg2:*
56: gb_hg3:*
57: gb_hg4:*
58: gb_hg5:*
59: gb_hg6:*
60: gb_hg7:*
61: gb_hg8:*
62: gb_hg9:*
63: gb_hg10:*
64: gb_hg11:*
65: gb_hg12:*
66: gb_hg13:*
67: gb_hg14:*
68: gb_hg15:*
69: gb_hg16:*
70: gb_hg17:*
71: gb_hg18:*
72: gb_hg19:*
73: gb_hg20:*
74: gb_hg21:*
75: gb_hg22:*
76: gb_hg23:*
77: gb_hg24:*
78: gb_hg25:*
79: gb_hg26:*
80: gb_hg27:*
81: gb_hg28:*
82: gb_hg29:*
83: em_hg0:*
84: gb_hg24:*
85: gb_hg24:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1944.8	99.7	3335	10 AF089868	AF089868 Homo sapi
2	1933.6	99.2	2943	85 HUMMUC18A	M29277 Human Isola
3	1933.6	99.2	3306	85 HUMMUC18B	M28882 Human MUC18
4	1341.4	68.8	2106	11 AB035506	AB035506 Rattus no
5	1333	68.4	2890	11 AB035508	AB035508 Mus muscu
6	1204.2	61.8	1988	11 AB035507	AB035507 Rattus no
7	1197.8	61.4	2772	11 AB035509	AB035509 Mus muscu
8	275.4	14.1	2106	4 CGHEMCAMC	Y08855 G.gallus mr
9	271.8	13.9	2584	4 CHK1GP	D49849 Chicken mrn
10	259.2	13.3	2015	53 HSMUC184	X68267 H.sapiens M
11	243.2	12.5	173038	76 HSMUC184	AP001003 Homo sapi
12	234	12.0	1998	4 GGHMCAMC	Y08856 G.gallus mr
13	232.4	11.9	2075	4 GGHMCAMC	Y08854 G.gallus mr
14	232	11.9	2365	4 CHKSGP	D38559 Chicken mrn
15	227.8	11.7	177595	76 AP001557	AP001557 Homo sapi
16	227.6	11.7	173038	76 AP001003	AP001003 Homo sapi
17	226	11.6	198902	76 AP001156	U89328 Bos taurus
18	222.6	11.4	317	3 BTMCA1	AP001156 Homo sapi
19	222.2	11.4	267	11 MM0297451	U89328 Bos taurus
20	213.8	11.0	270	53 HSMUC182	AJ297451 Mus muscu
21	180	9.2	198902	76 AP001156	X68265 H.sapiens M

BEST AVAILABLE COPY

[illegible]

ALIGNMENTS

X68268 H. sapiens M
 AP001557 Homo sapi
 X74528 M. musculus
 X74627 M. musculus
 X68264 H. sapiens M
 X68270 H. sapiens M
 AI297452 Homo sapi
 X68266 H. sapiens M
 AI297450 M. muscu
 M76678 Chicken Ig
 X64301 G. gallus mR
 S63276 SC1-100 ka
 AF109160 Mus muscu
 AF035511 Mus muscu
 AF214507 Mus muscu
 I66494 Sequence 14
 X83425 H. sapiens L
 AC003653 Human Chr
 AC026645 Homo sap
 AI035912 Streptomy
 AL105510 R. rattus n
 X80026 H. sapiens B
 AF206633 Streptoc
 AL161416 Leishmani

RESULT	1
AF089868	3335 bp mRNA
LOCUS	PRI 04-MAR-1999
DEFINITION	Human sapiens cell surface glycoprotein P1H12 precursor, mRNA,
ACCESSION	AF089868
VERSION	complete cds.
KEYWORDS	AF089868.1 GI:4336423
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE	1 (bases 1 to 3335)
JOURNAL	Gul, L., Chang, L., Browne, P.V. and Heibel, R.P.
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 3335)
TITLE	Gul, L., Chang, L., Browne, P.V. and Heibel, R.P.
JOURNAL	Direct Submission
REFERENCE	UMHC, 420 Delaware Street SE, Minneapolis, MN 55455, USA
FEATURES	Location/Qualifiers
source	1..3335
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/cell_type="umbilical vein endothelial"
	/note="Primary culture"
	24..1964
	/note="P1H12 antigen"
	/codon_start=1
	/product="cell surface glycoprotein P1H12 precursor"
	/protein_id="AAD17799.1"
	/db_xref="GI:4336424"
	/translation="MGDPRLVCAFLAACCCPRVAGVPGAEQAPAEIVVEVSGTA
	LKGLSGSGCNSLHWDSVHKKRRLTFEVRGOGSGEGEYERLSLDKGAITLA
	LTQVPRDERFLFCGKKRPSQEVRIORRYKAPFPFNINPLGIFVNSKEEVAAT
	CGRNGYPIPOVIWYKNGRLPEKRNRVHDSSTVSSGLYTLISLKKQIKEDKD
	AGPELCNVRLEPSGNHMKSEEVYVPVPEYKELVPEKPEGMKKEEDRYEIRCLDQ
	NPPHFSISKQPSSTREAEEETITDNGVLVTPARKHESGYECQGLDITMISLSE
	PEELIVNVSDVRSVPAPEREQSSLTITCEASSDLEFQMLRETDVLEGPVL
	QCHDILKREAGGGRCAVSPISPLRTQLNVNAIFGPMAFKERVMKENVNL
	SEASGSHPRPTSMKNNGTSAPDQDQSTAVSTLVAIVTPELTGVECTASDICKN
	VTISLELVNLTTLTPDSNTTGLSTSPASHATRANSSTERKLPEDESRGVIVAVIT
	VTILVAVIGAVLYFLYKGRKLPCKRSKQETILPSPKSSLVYEVMSDKLPEEMGLL

BASE COUNT	809 a	921 c	941 g	664 t	QSSGDKRAPDQCEKTYIDLRH"
ORIGIN					
Query Match	99.7%; Score 1944.8; DB 10; Length 3335;				
Best Local Similarity	99.9%; Pred. No. 0;				
Matches 1946: Conservative	0; Mismatches 2; Indels 0; Gaps				
OY	1	atggagcttccagagctgtgtctgtcgccctctgtccgcgcgcctgtcctgtcctgtccgc	60		
DB	24	ATGGGGCTTCCACAGCGTGTCTGCGCTTCTGTCTGTCTGCGCCCTGCTGCTGTCTGCTCGC	83		
OY	61	gtcagaggtgtgtcccgagagagctctgacagcctgtgccttgagcttgtgtgaggtgaaagt	120		
DB	84	GTCGGGGGTGTGCCCGGAGAGGCTGAGCAGCCGTGCGCCGTGAGCTGTGTGAAGTGAAGTG	143		
OY	121	ggcagacagacgctcttctgaaagtgtggcctctccagttcccaagagaacctgaacatgtc	180		
DB	144	GGCAGACAGCGCTTCTGAAAGTGGGGCTCTCCAGTCCCAAGCAACCTTAAGCAATGTC	203		
OY	181	ggctgtgttctctgtcccaagagaaagcgaagcgtcaatcttccgtgtgtgcgaagggccag	240		
DB	204	GACTGCTTTCTGTCTCCAAAGAGAAACCGAGCCTCATCTTCCGTGTCGCGCAGGGCAG	263		
OY	241	ggccagagcgaacctgtggagttacgacagcgtctagcctccaagaagaagagggactact	300		
DB	264	GGCCAGAGGGAACCTGGGGGATPAGACACCGGCTCAAGCCTCAAGCAGAGAGGGCTACT	323 *		
OY	301	ctggcccttactcaagtcaaccccccaagacgaagcgaatcttctgtgtccagggcaagcgc	360		
DB	324	CTGGCCCTTACTCAAGTCAACCCCCCAAGAGAGCGCATCTTGTGTGCCAGGGCAAGCGC	383		
OY	361	cctcgtgtccagagttaccgcatccagctccggtcttlaaagcttcccgaaagagctaac	420		
DB	384	CTCGGTCCCAAGAGTACCGGATCCAGCTCCGGGTCTACAAAGCTTCGGAGAGCAAC	443		
OY	421	atccagatcaacccccctgtgacatccctctgttaaacagtlaaagagagcttgaagagt	480		
DB	444	ATCCAGGTCAACCCCTGGGCACTCCCTGTGACAGTAAAGAGCCTGAGAGGCTGCTTACC	503		
OY	481	tgtgtgagaggaacgggtgatacccaattctctaagtctctgttaacaagaatgtgcgcgct	540		
DB	504	TGTGTGTGGAGGAGCAACGGGTATCCCAATCTTAAGTCAATCTGTGTGTAAGAAATGGCGGCT	563		
OY	541	ctgaagagagagaagaacccggttccacaattcagtcgtccagaagcvtgtgagttcgatgtgt	600		
DB	564	CTGAAGAGAGAAAGAACCGGGTCCACATTCAGTTCGCCAGACTGTGGAGTGAATGAT	623		
OY	601	ttgtacaacctgtcagaagtcttctgaaggcagaagcttgtttlaaagaagaacaaga	660		
DB	624	TTGTACACCTTGGAGAGTATTCTGAAAGGACACGCTGGTTAAAGAAAGCAAGATGGCCAG	683		
OY	661	ttttaacgttagcttcaactaacccggtctgcacagtgtgaaaccaatgaagagttccaaaggaa	720		
DB	684	TTTTTACTGTGAGCTCAACTAACGGGTCGCCAATGTGGAACCAATGAAGAGAGTCCAGGGAA	743		
OY	721	gtcacgctccctgttttcttaccgcgaagaagaaagtggtgtgtgaaagtgtggcccggtgtgaa	780		
DB	744	GTCACCTCTCCCTGTTTTTCTAACCCGAGAGAAAAGTGTGGCTGGAAGTGGACCCGTGGGA	803		
OY	781	atgtctgaagaaagagagccggtgtgaaatccaggtgttctgtgtgtatgtggcgaacctcaca	840		
DB	804	ATGCTGANGAGAGGGGAGCCCGGTGGAATTCAGGTGTTGGCTGTGATGGCAACCTCCACCA	863		
OY	841	caattcagatcatgcaagcagaaccccaagcaccaggggggagaggaagagaacaacaac	900		
DB	864	CACCTCACACATCACCAAGCAAGCAACCCACACCCAGGAGGCGAGAGAGACACCAACAC	923		
OY	901	gacacaggggtcctgtgtctgtgagcctgtccggaaagaaacacagtgtggcgtatgaa	960		
DB	924	GACACAGGGGGTCTGTGTGTGTGAGACCTGCCCGGAAGAAACACAGTGGCGCTATGAAATGT	983		

LOCUS	DEFINITION	HUMANUC18A	2943 bp	mrna	PRI	16-AUG-1994
QY	961	caggccctcgagacttgagacacatgatatcgtcctctagtgtaaccacaggaactctacgtgtg				1020
Db	984	CAGGCGCTCGACTTGGACACCACTGATATATCCCTGCTGAGTGAATACCAAGCACTACTGTTG				1043
QY	1021	aactcgtgtctcagcgcgcagagtgagtcgccgcagccctctagagacaggaagtcaagcgc				1080
Db	1044	AACATATGTGTGTGACGTCGCCAGTGTAGTCCGACGCCCTGTAGAGACAGAAAGGCACACAC				1103
QY	1081	ctcaacctgacactgtgtgagcagagagagtagccaagacctcagttccagttgctgtagagaa				1140
Db	1104	CTCACCTCGACCTGTGTGAGGACAGAGATGAGCAGAGACTCGATTCAGTTCAGTGTGAGAGAA				1163
QY	1141	gagacagcgccaggtgtcgtggaagaaggggcccgtgctcttaagttgatgatgacctgaaacggag				1200
Db	1164	GAGACACACACAGGTGTGTGGAAAGGGGGGCTGTGCTCTTCAATTGATCATCACTGAAACGGGAG				1223
QY	1201	gcaagagcgagcctacgcgcgtgagcgctcgtgtgccccagcatacccgagctgaacgcgcaca				1260
Db	1224	GCAAGAGCGGCTATCCCTGCGTGGCTCTGTGCCCCAGCATACCCTGGCTCGAACCCTCACTA				1283
QY	1261	caagctgtgtcacaagctgtgcccatttttggcccccttgatgtgcatctcaagagagagaaagttg				1320
Db	1284	CAGCTGTGCAACGCGGCCATTTTGGCCCCCTTGATGTGATTCATCAAGAGAGAAAGGTG				1343
QY	1321	tggtgtgaaagagataatgtgtgttgatcgtctctgtgaaagcgtcaggtgcaacccccggccc				1380
Db	1344	TGGTGTGAAAGGAATATGTGTGTAATCTTCTTGTGAACGTCGACGGCACCCCGGCCCC				1403
QY	1381	accctccctctggaacgctcaacgagcagagcgaagtgaacaagaaccaagatccacaagcgagtc				1440
Db	1404	ACCATCTCTGTGAACGTCGAACGGCAGCGGCAAGTGAACAAACCAAGATCCACAGCGAATC				1463
QY	1441	ctgagcaacctgaaatgtcctcgttgaccgccggagcgtgtgttgagacaggtgtgtgaatgacg				1500
Db	1464	CTGAGCAACCCCTGAATGTGTCGTGTCGACCCCGAGCGTGTGTGAGACAGGTGTGAATGCACG				1523
QY	1501	gctctcaacgacgctcgtggcgaaaaacaaccagatcctctctctgtagctgtgtcaattaaacc				1560
Db	1524	GCTCTCAACGACCTGTGGGCAAAAACACAGCATCCTCTTCTGTGAGCGTGTGCAATTTAACCC				1583
QY	1561	acctctacaaccagactccaacaacaacacgacctgaagccttcacatgcctgcaagctccat				1620
Db	1584	ACCTCTCACACCACTCCACACACACACTGTGCTTACGACTTCTCACATGTCGAATCTCTCAT				1643
QY	1621	accagagaccacaacgacacctccacaagagagagaagctgtccgagacccggagacccgggctgtg				1680
Db	1644	ACCAAGAGCCAAACACACACTCTCCACAGAGAAGAAAGTGGCGGAGCGCGAGACCGGGGGGTG				1703
QY	1681	gtcaatcgtgtgctgtgtgtgtgtgtcattccgtgtccctgaggggtgctgtgagcgctgtcctcat				1740
Db	1704	GTCATCGTGGGTGTGATTTGTGTGATCCTGTGCTCTGTGGCGGTGGCGCTGTCTCTCTAT				1763
QY	1741	ttctctctataaagaggaagcctgcccgtgtgaagcggtcccaaggagacagagatccaagctg				1800
Db	1764	TTCTCTTATTAAGAAAGGCAAGCTGCCGTGCAAGGGGCTCAAGGAAGCAGGGATCATCGCTG				1823
QY	1801	cccccgctctcgtlaaagcggaactgtgtgtgaagttlaagtcagaataagctccccaagaag				1860
Db	1824	CCCCCGTCTGCTAAGACGAACCTGTAGTTGAAGTTAAGTCACATTAAGTCCTCCAGAAAG				1883
QY	1861	atggagcctcctctcagggcagcagcggtgtgacaagaagggctcccgagagaccaaggagagaa				1920
Db	1884	ATGGGCGCTCTCGCGAGGGGACGACGGGTGACCAAGAGGGCTCCGGGAGACCCAGGAGAGAAA				1943
QY	1921	tacatgatctgtagcattagccccgaa	1948			
Db	1944	TTCATTCGATCTGAGGCAATTAGCCCCGAA	1971			

ACCESSION	M23277	cds.
VERSION	M23277.1	GI:530047
KEYWORDS	MUC18 glycoprotein; cell adhesion molecule; immunoglobulin-like sequence; integral membrane glycoprotein.	
SOURCE	Human cDNA to mRNA.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
AUTHORS	1 (bases 1 to 2943) Lehmann,J.M., Riettmuller,G. and Johnson,J.P.	
TITLE	MUC18, a marker of tumor progression in human melanoma, shows sequence similarity to the neural cell adhesion molecules of the immunoglobulin superfamily	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 86 (24), 9891-9895 (1989)	
MEDLINE	90099368	
AUTHORS	2 (bases 1 to 2943) Sers,C., Kirsch,K., Rothbacher,U., Riettmuller,G. and Johnson,J.P.	
TITLE	Genomic organization of the melanoma-associated glycoprotein MUC18: implications for the evolution of the immunoglobulin domains	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90 (18), 8514-8518 (1993)	
MEDLINE	93391884	
AUTHORS	3 (bases 1 to 2943) Johnson,J.P.	
TITLE	Direct Submission	
REFERENCE	Submitted (06-Oct-1989) Judith P. Johnson, Institute for Immunology, University of Munich, Goethestrasse 31, Munich, Germany 80336	
COMMENT	On Aug 16, 1994 this sequence version replaced gi:188859.	
FEATURES		
source	Location/Qualifiers 1..2943 /organism="Homo sapiens" /isolate="Juso" /db_xref="taxon:9606" /tissue_type="melanoma" /cell_line="Mel Juso" /clone="drop4.7" 8..1948 /codon_start=1 /product="MUC18 glycoprotein" /protein_id="AA20824.1" /db_xref="GI:530048" /translation="MGLPRLVCAFLAACCCPRVAGVPCGEADQAPAEVVEYEGSTRA LAKGSLQSGOGLSHVDMFVHKERKRLIFRRVGGQSGSEGEYQRLSLQDRCATLA LTQVTPODERFLFCQGRPRSQEYRIOLRYKAAPEEINIVNPGLIVNSKEPEEAT CVGNQYPIPOVIVYKNGRPLKEKNKHVHIOSSQTVSSGTYLTSLKOLVEDKD AQPCEINLRPLPSGNHMKSEYVIVPYFPEKYMFLVEYVGMKEEDRYEIRCLAG NPPHESTIKONPSTREAREEETTDNKLVLLEPAKKHSGRYEGQANNITDMLISLSE PQELLVIVSDVRVSPAAPEQEGSSLTLTCEASSODLEQWUREETDYLIEGPVL QLHDKREAGGVCRCVASVPSIPGLNFTOLVKLAIFGPMAAFKERVYVAKENNVNLM SCENSGRPRTISWNVAGTASEODQDQRYSLTNLVLPFELLETVGECAASNDLGNV TSLIFLELVNLTTLTPDSNTTGTSTSTASPHRTANSTERKLPEESGGVIVAVIV VCLIVLVAGVILVLETKKKGKLPKRRSGKEITLPPSRKTEIVVEVKSQDLPEMGLL QGSGDKRAREDDQSEKTYIDLRH" 8..91 /product="MUC18 glycoprotein signal peptide" 92..1945 /product="MUC18 glycoprotein" 122..373 /note="immunoglobulin-like (V set) domain I" 467..694 /note="immunoglobulin-like (V set) domain II" 795..988 /note="immunoglobulin-like (C2 set) domain" 1076..1249 /note="immunoglobulin-like (C2 set) domain" 1337..1525 /note="immunoglobulin-like (C2 set) domain" BASE COUNT 668 a 863 c 858 g 554 t	
ORIGIN		
Query Match	99.2%; Score 1933.6; DB 85; Length 2943;	

Best Local Similarity 99.5%; Pred. No. 0;
Matches 1939; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 1 atggagctccagagctgtctgcgctcttctgcgcgctgtctgtctgtcctcgc 60
|||||
Db 8 ATGGAGCTCCAGAGCTGTCTGCGCTTCTTCTGCGCCGCTGCTGTCTCTCCG 67
OY 61 gtgcgggtgtgtccggagaggtgtgacagctgcgcctgagctgtgtgagtgag 120
|||||
Db 68 GTGCGGGGTGTGCTCCGAGAGGCTGAGCAGCTGCGCTGAGCTGTGTGAGGTG 127
OY 121 ggcagcaagccccctctgaagtgagcctctcccaagtcaccaagcaacctcag 180
|||||
Db 128 GGCAGCAGACGCTTCTGAAAGTGGGCTCTCCACTCCCAAGGCAACTCAGCCAT 187
OY 181 gactgttctctgtcccaagaagaagcggagctcaactctccgtgtgcgcaagg 240
|||||
Db 188 GACTGTGTTTCTGTCCCAAGAGGAGGAGCGACCTCATCTTCCGTGCGCCAG 247
OY 241 ggcagagcgaacctgtggagatgacagcagcgtcagctccaggtacagagggt 300
|||||
Db 248 GGCAGAGCAACCTTGAGAGTACGAGCAGGCTCAGCTCCAGAGCAGAGGGCTACT 307
OY 301 ctggcccttgaactaaatcaaccccccaagaagcagcagctctctgtgcagagg 360
|||||
Db 308 CTGGCCCTGACTCAAGTCAACCCCCCAAGAGAGCGCATCTTGTGTGCCAGGG 367
OY 361 cctcgtgtcccaagagatcagcagatcagctccgcgtctacaagaatccgagag 420
|||||
Db 368 CCTCGGTGCCAGAGATACCGCATCCAGCTCCGCTGTACAAAGCTCCGAGAGCC 427
OY 421 atccaggtcaacccccctgtgacatccctgtgaaagtaaggaagcctgaagag 480
|||||
Db 428 ATCCAGGTCAACCCCTTGCGCATCTTGTGAACAGTAAGAGCTTGAGAGGTG 487
OY 481 tctgtgagaggaagcggatcaacccatctcctcaagtcattctgtgtacaaga 540
|||||
Db 488 TGTGTAGGAGGAAGGAGTACCCCATTCCTCAAGTCAATGTGGTACAAAGAT 547
OY 541 ctgaagaggaagaagaacgggttcaacattcaagtcgttcccaagctgtgag 600
|||||
Db 548 CTGAAGAGAGGAAGAACCGGGTCCACATTCAGTCTGTCCAGACTGTGAGTGT 607
OY 601 ttgtacacctgtgagatattctgaagcagacagctgtgttaagaagaacaaga 660
|||||
Db 608 TTGTACACCTTGAGAGTATTCTGAAGGCACACCTGTGTTAAAGAGACAAAG 667
OY 661 ttttactgtgagctcaactacccgctgcgcagctgaggaaacacatgaaggag 720
|||||
Db 668 TTTTACTGTGAGCTCAACTAACCGGCTGCCAGTGGGAGACACATGAGAGT 727
OY 721 gtcaacgctcctgttttctacccagagaagaagtgtgtgtgagagtcgagc 780
|||||
Db 728 GTTACCGTCCCTGTCTTTTACCCGACAGAAAAGTGTGTGTAATGTGAGCC 787
OY 781 atgctaaaggaagggagcggctgtgaatcaggtgttctgtgtgtgtgtgtgt 840
|||||
Db 788 ATGCTAAGGAAGGAGGAGCCGCTGGAATCAGGTGTTGGCTGTATGCAACCT 847
OY 841 cacttcaagatcagcaagaagacccccagacccaagggagagcagagaagaaca 900
|||||
Db 848 CACTTCAGCATCAGCAAGCAGAACCCACAGCACAGAGGAGGAGAGAGAACAC 907
OY 901 gataaaggggtctctgtgtctgtgagcctgcgggaagaacacagtgtgagat 960
|||||
Db 908 GAAACAGGGGTCTCTGTGTGTGAGCTGCCGGAAGAACAGTGGCGCATGATGT 967
OY 961 cagggtctggaactgtgaacacatgatatacgtctgtgtgtgtgtgtgtgtgt 1020
|||||
Db 968 CAGGCTGGAACCTTGACACCATGATATGCTGTGTGTGTGTGTGTGTGTGTGT 1027
OY 1021 aactatgttctgtgagctccgagtgagttcccgagccccgtgaagaagacagga 1080
|||||

Db 1028 AACTATGTTGCTGAGCTGCCAGTGTAGTCCCGACGCCCTGTGAGACAGAAAGCAGCAGC 1087
OY 1081 ctcaacctgaacctgtgtgagcagaagatagccaggacctcgagttccagtggtgtgagaa 1140
|||||
Db 1088 CTCACCTGTGACTGTGAGAGCAGAGAGTATGACAGACCTCGAGTTCCAGTGTGCTGAGAGAA 1147
OY 1141 gagacagcagcgtgtgtgtgaagaaggagctgtgttcaagttgtgaatgaatgaacaggag 1200
|||||
Db 1148 GAGACGAGACAGGTGTGTGAAGAGGGGCTGTGTCTTCAAGTTGATGACCTGTGAAGGGGAG 1207
OY 1201 gcaagagcggcctatcgcgt 1260
Db 1208 GCAGAGGCGGGCTATTCGT 1267
OY 1261 cagctgtgtcaagctgt 1320
Db 1268 CAGCTGTGTCAAGCTGT 1327
OY 1321 tgggtgaagaagaatgt 1380
Db 1328 TGGGTGAAGAGAAATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1387
OY 1381 accatctctgtgaacgttcaacgagcagcgtgtgaacaagaaccaagatccacagcagatc 1440
Db 1388 AACATCTCTCTGGAAGCTCAACGCGCAGGCAAGTGAACCAAGATCCACAGCGATC 1447
OY 1441 ctgagcaacctgt 1500
Db 1448 CTGAGCACCTGTGAATGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1507
OY 1501 gcttcccaagcagcctgt 1560
Db 1508 GCTCTCAACGACACTGT 1567
OY 1561 acccttcaacacagacttccac 1620
Db 1568 ACCCTCACACACAGACTCCAC 1627
OY 1621 accagagccaacacagaccttccacagagagaagaagctgtccggaagccggagagccgg 1680
Db 1628 ACCAGAGCCAACAGCAGCTCCACAGAGAGAAAGTGTGCGGAGCGGAGACCGGGGCGTG 1687
OY 1681 gtcaatcgt 1740
Db 1688 GTCAATCGT 1747
OY 1741 ttcctctataagaaggagcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1800
Db 1748 TTCTCTATTAAGAGGGCAGACCTCCGTGCGAGCGCTCAGGAGGAGGAGATCACGCTG 1807
OY 1801 ccccgctctgttaagagagaaactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1860
Db 1808 CCCCCGTCTGTGAAGCCAGAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1867
OY 1861 atggcctctctgtcagggagcagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1920
Db 1868 ATGGGCTCTGTGAGAGGAGCAGCGGTGTGACAAAGAGGGCTCCGGAGACCAAGGAGAGAAA 1927
OY 1921 taacatgatactgtgaagcattagccccgaa 1948
Db 1928 TACATCGATCTGAGGCAATTAGCCCCCGAA 1955

RESULT 3
LOCUS HUMMUC18B 3306 bp mRNA PRI 19-AUG-1994
DEFINITION Human MUC18 glycoprotein mRNA, complete cds.
ACCESSION M28882
VERSION M28882.1 GI:529723
KEYWORDS MUC18 glycoprotein; cell adhesion molecule; immunoglobulin-like
sequence; Integral membrane glycoprotein.
SOURCE Human cDNA to mRNA.
ORGANISM Homo sapiens

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 3306) Lehmann,J.M., Riethmuller,G. and Johnson,J.P. MUC18, a marker of tumor progression in human melanoma, shows sequence similarity to the neural cell adhesion molecules of the immunoglobulin superfamily
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 86 (24), 9891-9895 (1989)
REFERENCE	2 (bases 1757 to 1945) Sers,C., Kirsch,K., Rochbacher,U., Riethmuller,G. and Johnson,J.P. Genomic organization of the melanoma-associated glycoprotein MUC18: implications for the evolution of the immunoglobulin domains
MEPLINE	Proc. Natl. Acad. Sci. U.S.A. 90 (18), 8514-8518 (1993)
AUTHORS	93931384
REFERENCE	3 (bases 1 to 3306) Johnson,J.P. Direct Submission Submitted (06-OCT-1989) Judith P. Johnson, Institute for Immunology, University of Munich, Goethestrasse 31, Munich, Germany 80336
COMMENT	On Aug 15, 1994 this sequence version replaced gi:188861.
FEATURES	
Source	Location/Qualifiers 1..3306 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_line="Mel Juso" /tissue_type="melanoma" /clone="zapyl, drop 1, drop 4, drop 4.2, drop 4.7, drop 4.11"
sig_peptide	8..91 /product="MUC18 glycoprotein signal peptide"
CDS	8..1948 /codon_start=1 /protein_id="AA20922.1" /translation="MGFLRIACFLAACCCPPVAGVGFAEPAPELVEVEGSTA LTKGSGOGSLSHDWEVSFKKFTLIPIVRVGGOSGEDEQSLSDRGATTA LTGYPDREFILCGKRPSOEYRIOLRYAKAEPRIOYNPLGITVNSKEPEVAT CVGNNGPIPIGVITKKGRPLKEKNRKHIDSSOTVSSGLTTLSTLKQLVEDKD AQFCELNTRYLPSSNHHKESREVTVPVFPEEKWALEVEYGMLEKDDRAEIRCLAGE NPHPFSISKSPSTRAEETINDVLVLEPAKRBHSRGVECOANMLDTMSLSGE POELIKRVASDVRSPPAEREGSSITLICEASSODLEFORMLETDOVLERGPIV QLHLKREAGGVCVA SVSIPGLINTFOIKALIFGPMPAFERKYWKVENNVLMNL SCESGSRPTISNNVGTASEODORDYSITNTIVTPLEILETGVCASNDCKNK TSILFLVNTITLPTSNTTGTGSTASTASHPTRANSSTERRKIPEPSRKGVIAVAY VCIILVALICAVLYFLYKGRKLPCRSGKDETLPSPRTKTELVEVKSPDKLPBEMGILL QGSSGDKRAPEDQGEKYIDLHK"
misc_peptide	92..1945 /product="MUC18 glycoprotein"
misc_feature	122..373 /note="immunoglobulin-like (V set) domain I"
misc_feature	467..694 /note="immunoglobulin-like (V set) domain II"
misc_feature	793..988 /note="immunoglobulin-like (C2 set) domain"
misc_feature	1076..1249 /note="immunoglobulin-like (C2 set) domain"
misc_feature	1337..1525 /note="immunoglobulin-like (C2 set) domain"
misc_difference	replaced(1766,g')
misc_difference	/note="g is deleted"
misc_difference	/citation=[1]
misc_difference	1880*1881 /note="insert g"
polya_signal	/citation=[1] 3273..3278
BASE COUNT	799 a 916 c 924 g 667 t
ORIGIN	

[illegible]

Dh	1028	AACATNTGTGTACGTCGAGAGTACCTCCCGACGCCCTTGAGAGACAGAAAGCAGAC	1087
Qy	1081	ctcaaccctgaaccctgtgagagcagagatgaaccagacctcgaatcccaatgctctgagaa	1140
Dh	1088	CTCACCTCTGACCTGTGTAGCGACGAGACTAACCCAGACCTTCGAGTTCCAGTGGCTGAGAGAA	1147
Qy	1141	gagacagagccagagtgctcgtgaaagggcccttgatcgaatgtcatgacctgaaagcgag	1200
Dh	1148	GAGACACACACAGGCTGCGGAAAGGGGGCCCTGTGCTTCACTTGCATGACCTGAAAGGGAG	1207
Qy	1201	gcaagagcgagcatcgtcgtcgtgagctctgtgcccagataccggagctgaaccgcaca	1266
Dh	1208	GCAAGAGCGCGCTATCCCTCCGTGGCTGTGTGCCAGATACCGCGCTGAAACCGACAA	1267
Qy	1261	caagctgtgtcaacgttggtcccaatlttttgccccctctgtagtgcatlcaagagagaaagt	1320
Dh	1268	CAGCTGGTCAAGCTGGGCCATTTTGTGGCCCCCTTGATGGCATTCAGAGAGAGAAAGTG	1327
Qy	1321	tgggtgaagagaatagtgtgttgaatctgtctctgttgaagctcagggcaaccccgccc	1380
Dh	1328	TGGGTGAAGAGAAATATGGTGTGAATCTGCTTTGGAAGCCCTCAGGGCACCCCGGCC	1387
Qy	1381	accaatccctccggagactgcacacggcagagaaagtgaacaaagaccaaagatccaagagatc	1440
Dh	1388	ACCATCTCTCTGGAAACGCTCAACGGCACGGCAAGGAACAAGACCAAGATCCACAGCAGTC	1447
Qy	1441	ctgagacacccctgaatgtctctctgtgaccgccgagactgttgaagacaggtgttgaatgcacg	1500
Dh	1448	CTGAGACACCTGAAATGTCTGTGACCCCGAGACTTGTGAAACAGGTGTTAAATGACAG	1507
Qy	1501	gctctcaagacacttggtgcaaaaacaccagacatccctcttccttgagactgtgtcaatlaac	1560
Dh	1508	GCTCTCAACGACCTGGGCAAAAACACAGATCTCTTCTGGAGCTGTGCAATTTAAC	1567
Qy	1561	accctcaacacagactctcaacaaacaaacactgtgactaagactlccacgtccagctccat	1620
Dh	1568	ACCTCTCACACCACTCTCAACACACACACTGGCTTAGCACTTCCACTGCCAGTCTTAT	1627
Qy	1621	accagagaccaaacacactccacacagagagaaagctltgcggaacccggaagccggggcgt	1680
Dh	1628	ACCAAGACCAACACGCACTCTCACAGAGAAAGCTCCGGACCCGAGAGCCGGGGCGG	1687
Qy	1681	gtcaatcgttgagctgtgaatgtgtgtcaatccctgtgtccctgtgcggtgcgtggcgctgtccctat	1740
Dh	1688	GTCATCTGTGGCTGTGATTTGTGTCAATCTGTGTCCTGTGGCGCTGTCTGTCTTAT	1747
Qy	1741	ttctctcataaagagggcagaagctgcgcgtgtgaagcggtctcaaggaaacagaaataatcacgtc	1800
Dh	1748	TTCTCTTATTAAGAAAGGCAAGCTGCCGTGAGGCGCTCAAGGAAGCAGAGATCACGCTG	1807
Qy	1801	ccccgctctcgtlaagagcgaactgttagtltgaagtlaagtlcaagataaagctccacagaag	1866
Dh	1808	CCCCCGTCTGTAAAGACCGAATTTGATGTTGAAGTTAAGTCAGATTAAGTCTCCAGAAAG	1867
Qy	1861	atggagctctcgtcaagggcagcagcgtgtgacaagaagggctccggtgagaacagagagaagaa	1920
Dh	1868	ATGGGCTCTCTCGACGGGCAACAGCGGTGACAAAGAGGGCTCCGGGAGACCAAGGAGAGAAA	1927
Qy	1921	taaatgatatctgaaggaatttagccccaag 1948	
Dh	1928	TACATTCGATCTGAGGCAATTAAGCCCCGAA 1955	

RESULT	4
LOCUS	AB035506
DEFINITION	AB035506 2106 bp mRNA
VERSION	AB035506.1
KEYWORDS	GI:10566948
SOURCE	1-glycerin/MUC18
ORGANISM	Rattus norvegicus (strain:Sprague-Dawley) male cDNA to mRNA.
	Rattus norvegicus
	Eukaryota; Metacoa; Chordata; Craniata; Vertebrata; Euteleostomi;

	REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	AUTHORS	Rattus.
	TITLE	1 (bases 1 to 2106)
	JOURNAL	Taira,E., Okumura,S. and Miki,N
	REFERENCE	Functional analysis of rat gicerin/MUC18
	AUTHORS	Published Only In Database (2000) In press
	TITLE	2 (bases 1 to 2106)
	JOURNAL	Taira,E., Okumura,S. and Miki,N.
	FEATURES	Direct Submission
	SOURCE	Submitted (03-DEC-1999) to the DDBJ/EMBL/GenBank databases. Eichi Taira, Osaka University Medical School, Osaka 565-0871, Japan (E-mail:eclair@pharm.med.osaka-u.ac.jp), Tel:+81-6-6879-3521, Fax:+81-6-6879-3521)
	ORIGIN	Location/Qualifiers
	CDS	1..2106
	gene	/organism="Rattus norvegicus"
		/strain="Sprague-Dawley"
		/db_xref="taxon:10116"
		/sex="male"
		17..1963
		/gene="1-gicerin/MUC18"
		17..1963
		/gene="1-gicerin/MUC18"
		/product="1-gicerin/MUC18"
		/codon_start=1
		/protein_id="PABI6048.1"
		/db_xref="GI:1056949"
		/translation="MCLPLVCAFLFAACCCRSATGVPEGKOPTPPDPPEVEGN TALSCGAHPSGNFSOVENFLLHKRPQIPPIRVHQKGSEPGTEHRSLHPGAT LALSQVPDHDRMLCKSKOPRODHYVOLQYKAPEPTIOANYLGIVDIOLKEV ATCGRNQYPIPOIWIYNKRNPLOEEENRHIOSSQVESSLYLTKLSVARKEED KDAPGYCEVSYPSPGNRMKESKEVIPLYAEKVWVEEVGLKEGDHVKTIRCTL DNPLEHTTIINKNPNSTEMEESTDENLLISLEPAKHHSVCOSLDLETVMLS SDPLEILVNYSVDQVPRTAPEVEDGSLLTCRKSENDDLEEFLRRKTOLLGKGP TLQINNYREAGSGRYLCVAASPVRPLNTPRYSVGITGGSPMMAKKKKWAAENAML NLSCASHHOPTYSMHINSATSATEKNMPDYVTSLNIANLYPELLETAEETANSNG SYTVIIILIKTTLTTHDSSOTTGLSTPVSHSHANSISTEKLLPDGEAGVIYA VLICLVLAVALGATLIYYFYFKRLPCGRSGKOETLLPTRISEVFVEVKSOKLPDEMA LLOGSGDKRAFGOGGEKYIDLNH"
	BASE COUNT	536 a 573 c 587 g 410 t
	Matches 1577; Conservative 0; Mismatches 366; Indels 6; Gaps 1;	
Query Match	Best Local Similarity	68.8%; Score 1341.4; DB 11; Length 2106; 80.9%; Pred. No. 9.8e-282;
Dn	1 atgggagctcccgaggatgtcgttcgcaccttctgctgcgcgcctcagtcatctgactcgc	60
Db	17 ATGGGGCTGCCCAGGTGTGTCGTGCACCCTTTATTATTCGCTGCTGCTGCTGCTGCAC	76
Qy	61 gtccgagtgtagccccgaagaagctltgagca-----cttcgcacgttcgtagtgaaagt	114
Db	77 GCCACGGGGTGTCGACAGAAGAGGAMAAGCACGCCAACCCCAGCGCGGACCGGGTAGA	136
Qy	115 gaagttgggaagacaagaccctcttgaagtgtgggcctctcccacgtgcccaaaggccaacccag	174
Db	137 GAAGTGGGACAACGGCCCTTCTGAAGTgttg6CCCCCGCATCTCCCTTCAGGACTTCCAC	196
Qy	175 catgtcgaactgattctctgtccacaagaagagaagagcacgtcalctctcgttgtgctcag	234
Db	197 CAGGTGGAAATGTTTTTGAATTACAAAAGGAGGAGATACCACATTTTCCGTGTGCACAG	256
Qy	235 ggccaagggccaagaagcaactctgggaagttaagaagaagcagcgctctcagctccagaagcaagag	294
Db	257 GGCAAAGGCCACAGTAACCTTGGAATAATATAACACCGCTCCACCTCCACCGGCCACAGA	316
Qy	295 gctaacttgacctgaactcaagtaaccccccaagaagcagcatctcttgtgtccaggc	354
Db	317 GTACTCTGTGGCCCTGATGTCAAATCACTCCCATATGATGACCGAAGTTCCCTGTAGAGAC	376
Qy	355 aaagcactcgtgtcccaagagataccgatccagctcagctcgtctaccaaatgctccgaagag	414

```

Db 377 AAGCAGCCACGGCCACGATGATGTTAGTCTGACGCTGACAGCTTACAGGCTCCAGAGAG 436
Oy 415 ccaaacatccagtgtaaacccccctggcattccctgtgaagaatgaagagctgaagatc 474
Db 437 CCAACTATCCAGGCAAGGCTGCTGGGCATCCACGTTGAGATCAAGAGCTTAAAGGAGT 496
Oy 475 gctacctgtgtaggaggaacggtgtaacccattctcctaagatcatcgtgtacaagaatg 534
Db 497 GCTACTGCTGTGGGAGGAGATGCTACCCCATCTCTCAAGTATCTGGTACAGATGAGC 556
Oy 555 cggcctctgaagaggaagaagaaccggttccacattcagtcgltccacagctgtgagtcg 594
Db 557 CGGCCCTTGCAAGAGAGAGAGAACCGGCTTCATATCCAGTCTGACAGACCGTGGAGTCC 616
Oy 595 agtgtgtgtacacctgtcagagatcttctgaagcagacagctgtgttaagaagaagaat 654
Db 617 AGTGCTTGTACCTTGAAGAGGAGTGTCTAAGTGCACGGGTAGTCAAGGAGACAAAGAT 676
Oy 655 gccagtttactgtgagctcaactaccgctgcagctgagtggaacacatagaagatcc 714
Db 677 GCCCAGTTTACTGTGAGAGTCAAGTACCGGCGCCGACGAGGGAACCGCATGAAGATCT 736
Oy 715 aggaagatcaacgcttccctgttcttctacccagagaagaagtgctgagtgagtc 774
Db 737 AAGGAGGTCACTGCTCCCTGCTTATACCTCGAGAAAAAGTGTGGGTGAGGTAGAACCC 796
Oy 775 gtggaagatcgtgaagagggagccgctgtggaatcagtgctgtgtgagtgagtc 834
Db 797 GTGGAGTGTCTGAAGAGGAGGAGCCATGTAAGATGAGTGTCTAAGTACGCGCAACCT 856
Oy 835 ccaaccacatcgaatcagcaagaagacagaccacagcaagaagagaggaagaagaca 894
Db 857 CAACCCCACTTACATCAACAAGAGAAACCCACACACTGAGAGATGGAAGAAGAGAGAC 916
Oy 885 accaagcaacaggggtccctgtgtctgagcctgcgcgaagaagaacagtgagtc 954
Db 917 ACCGATGAAAAAGGCTCTCTCTGCTTGGAGGCGCCGACAGAACACACAGTGGGCTTAC 976
Oy 955 gaatcagggcctgtgactggaacacatgatcgtcgtgtgagtgaggaacacagaacta 1014
Db 977 CAATCTCAGAGCTTGAAGCTTGGAAACTACAGTTAGCTGTGCAGAGTACCCCTTGAAGCT 1036
Oy 1015 ctgtgtaactatgtctcgaagctccgagtgagtcgagcagccctgagagaagaagaagc 1074
Db 1037 CTGGTAACATATGTCTGTGATTTCAAGTGAATCAACGCCCCGTAAGTCCAGAGAGT 1096
Oy 1075 agcagctcaccctgaacctgtgagcagagtagcagagaccctcgagttcagtgctg 1134
Db 1097 GATACCTTCACACTGACTGTAAAGCAGAGAGTAACACAGGACCTTGAGTTGAGTGGCTG 1156
Oy 1135 agaagaagagagagcagagtgctgtgaagaaggggctgtgtcattcattcagtaactgaa 1194
Db 1157 AGAGCAAGAGACGGCCACTTGTGGAAAGGGGCCCATCTTCACTTAACAACAGCTGAAA 1216
Oy 1195 cggagagcaggaagggagctatcgtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1254
Db 1217 AGGGAAGAGAGGGGGAAGATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1276
Oy 1255 cgcacagcctgtgtcaacgtgtgcaattttgtgccccctgtgagtgagcattcaagaag 1314
Db 1277 CGCACACGCGGGGTCAGGTGGGCAATTTTGGTCCCGTGGATGGCAGCAAGAAGAGAGG 1336
Oy 1315 aagtggtgtgtgaagaagaataatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1374
Db 1337 AAGGTGTGGGCGCAAGAAATGCTGAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1396
Oy 1375 cggccacacatcctcgtgaacgctcaacgagcaagcaagtgaaacagaacatccacag 1434
Db 1397 CAGCCACACATCTCTGTGAATATCAATGCTGTGGCACTGAATGGAACCTGTATCTCTCAG 1456
Oy 1435 cgaagtctgtgagacacctgaatgtctcgtgtgaccccgagctgtgtgtgtgtgtgtgt 1494

```

```

Db 1457 ACAGTAGTAGACACCTTGAATGCTTGTGACCCAGAGCTTCTGGAGACCGGTGCAGAA 1516
Oy 1495 tgaaggtctccaagacctgtggcaaaaacaccagcatctcttctgagctgtgtcaat 1554
Db 1517 TGTACAGCTCCACATCCCTGTGGCTTCATACACCAGCTCATCTATCTTGAAGCTGTCACT 1576
Oy 1555 ttaacacacctacacagacctcaaacacacacacacacacacacacacacacacacacac 1614
Db 1577 TTATACCATCCCTACACATATCTCTACACCAACCACTGCGCTGACACCCCAAGTCAGT 1636
Oy 1615 cctcatlacagagcaacagcactccacagagagaagaactgtgcgagccgagagagccg 1674
Db 1637 CCTCACAGAGAGCCACAGCAACCTCCACAGAGAAAGAACTGCTCCACAGCAGAGAGCAAA 1696
Oy 1675 ggcgtgtgtatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1734
Db 1697 GGTGTGCTATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1756
Oy 1735 ctctattctctataagaagagcaagctgtgcgtgtgagcgtctcagaggaagagatc 1794
Db 1757 CTCTATTACTTATACAGAGAGGCAATCTGCTTGTGACCTTCAGAGAAACAGAGATC 1816
Oy 1795 acgtgtccccgtctcgtlaagaagcgaactgtgagtgaaagtgaaagtgaaagtgaaag 1854
Db 1817 ACGTGTCCCGCCAGCTGTGAAGATGTAATTTGATTGAAGTTAAGTCAAGATTAAGCTCCA 1876
Oy 1855 gaaagatgtgctctcgtgagagcagcagcgtgtgacaagaaggtctccggaagcaagga 1914
Db 1877 GAAGAGATGCTCTCTCTTGAAGGAGCAACGGTGAACAGAGGCTCCAGAGACAGGGA 1936
Oy 1915 gagaatacatcgatctgtgagcattagcc 1943
Db 1937 GAGAAATACATGATCTGAGGATTAAGCC 1965

RESULT 5
AB035508
LOCUS AB035508 2890 bp mRNA ROD 03-OCT-2000
DEFINITION Mus musculus mRNA for 1-glicerol/MUC18, complete cds.
ACCESSION AB035508.1 GI:10566952
VERSION 1-glicerol/MUC18.
KEYWORDS Mus musculus (strain: BALB/c) male cDNA to mRNA.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 2890)
AUTHORS Taira, E., Okumura, S. and Miki, N.
TITLE mouse glicerol/MUC18
JOURNAL Published Only in Database (2000) In press
AUTHORS Taira, E., Okumura, S. and Miki, N.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) to the DDBJ/EMBL/GenBank databases. Eiichi Taira, Osaka University Medical School, Department of Pharmacology, 2-2 Yamada-oka, Suita, Osaka 565-0871, Japan (E-mail: etaira@pharm.med.osaka-u.ac.jp, Tel.: +81-6-6879-3521, Fax: +81-6-6879-3521)
FEATURES
source
1..2890
Location/Qualifiers
1..2890
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/sex="male"
20..1966
/gene="1-glicerol/MUC18"
20..1966
/gene="1-glicerol/MUC18"
20..1966
/codon_start=1
/product="1-glicerol/MUC18"
/protein_id="BAB16050.1"
/db_xref="GI:10566953"
/translation="MGIPKIVCFLEFAACCCBRAGVPEBEKOPVPTDIVEAVGS

```

[illegible]

QY	1915	gagaatacatcatcgtacgtgagcattag	1941
Db	1940	gagaatacatcattgattcgtgagcattag	1966
RESULT	6		
LOCUS	AB035507	1988 bp	03-OCT-2000
DEFINITION	Rattus norvegicus mRNA for s-gicerin/MUC18, complete cds.		
ACCESSION	AB035507		
VERSION	AB035507.1	GI:10566950	
KEYWORDS	s-gicerin/MUC18.		
SOURCE	Rattus norvegicus (strain:Sprague-Dawley) male heart cDNA to mRNA.		
ORGANISM	Rattus norvegicus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
AUTHORS	1 (bases 1 to 1988)		
TITLE	Taira, E., Okumura, S. and Miki, N.		
JOURNAL	rat s-gicerin/MUC18		
REFERENCE	Published Only in Database (2000) In press		
AUTHORS	2 (bases 1 to 1988)		
TITLE	Taira, E., Okumura, S. and Miki, N.		
JOURNAL	Direct Submission		
source	Submitted (03-DEC-1999) to the DDBJ/EMBL/GenBank databases. Eichl Taira, Osaka University Medical School, Department of Pharmacology, 2-2 Yamada-Oka, Suita, Osaka 565-0871, Japan (E-mail:etaira@pharm.med.osaka-u.ac.jp, Tel: +81-6-6879-3521, Fax: +81-6-6879-3521)		
FEATURES	Location/Qualifiers		
gene	1..1988		
CDS	/organism="Rattus norvegicus"		
	/strain="Sprague-Dawley"		
	/db_xref="taxon:10116"		
	/sex="male"		
	/lssue_type="heart"		
	17..1837		
	/gene="s-gicerin/MUC18"		
	17..1837		
	/codon_start=1		
	/product="s-gicerin/MUC18"		
	/protein_id="BAB16049.1"		
	/db_xref="GI:10566951"		
	/translation="MGLRLVCALFLPACCCRSATGVPGEKQPTPPDEVEVGN TALLCGAPHSAGNSQVEMFLIHKEQIDPFRHQGQSGEPEYHRSLHGPAT LALSGVPHDDRMFLKSKQRPQDHYVOLQVYKAPERTIOANVLGIHVDIQLEKEV ATCGRNGVPIPOVITWNGKRPLOEENRVAIOSQYVESGLTASVSARVKEK KDAQFCEVSRPSPSGNMRKSEKVTVPVLPARKVWVEVPGVLLKEGDHVTRICLT DGNOPHFVITINKKPNSTEMHEESTDENGILSTEPADKHHSGVYCOSLDLETTVMLS SDPELANTVSDVQVDPPIAEVQEGDSLTITCAESNODLEFELMDLTKQLLKGK ILQIANNNKREAGGRYLCVAVSPVSGLNRTFRSVSGIFSEPMNAEKRWAAQENAMI NLCSAGNHQPOPTITSMNINIGATSEMNDDPQVTVSLVNLVLPELLETAGECTASNSIG SYTTVILIKLTATLTLLIHDSQTTGLSPGVSPHSRANSTTEKKLPQGESKGVIVAV VIVTQVLAVLGAATLYFYFKKGRPLPCGSPSGNQEEMERNMTSI"		
BASE COUNT	503 a	544 c	553 g
ORIGIN	388 t		
Query Match	61.8%;	Score 1204.2;	DB 11; Length 1988;
Best Local Similarity	79.8%;	Pred. No. 6.9e-252;	
Matches 1435;	Conservative	0;	Mismatches 356; Indels 6; Gaps 1;
QY	1	atggagcttccacgagctgctgtcgccttctgtcgcgcgcctgctgtcgtcgtcgc	60
Db	17	ATGGGGCTGCCAGGCTGCTGCGCCCTTATTATTCGCTGCTGCTGCTGCTGCAGC	76
QY	61	gtcgcgagtgtagcccgagagagctgagagc-----ctcgcgctgaagctgtagtg	114
Db	77	GCCACGGGTGTGCCACGAGAGAAAGCACCACACCCACGCGCGGAGGAGGTA	136
QY	115	gaagtggcgagacacgaccttctgaagtgtagcctctccacagtcacaaagcaactcagc	174

Db	137	GAAAGTGGCCAAACACG6CCTTTCTAAGTGTGGCCCCGCACATCTTCAGGCAACTTCAAC	196
OY	175	caatgcagacttggtttcttcgaccacaagaagaacggcagctcatctccgtgtgcgcag	234
Db	197	CAGAATGGAAATGGTTTTTGATTTCACAAGAGAGGAGAGTAACCATTMTTCGTGTGCACAG	256
OY	235	ggccaaggccaaagcgaacctggggagtacagcacgctcgaagcctccaaagcaagaag	294
Db	257	GGCAAGGGGCCAGAGTAGAACCTCGGGGAATATGAAACACCGCCTCGACCTCCACGGCCAGGA	316
OY	295	gctactctcgccctcgaccctaagtlacccccaccaagaagcagtaetctcttctgtgccaggc	354
Db	317	GCTACTCTGGCCCTGAGTCAAAGTACATCCCATAATATCACCGAATGTTCTCTGTATAGAC	376
OY	355	aagcgccctcgctcccaagataccgatccagctccgcgtlacaagaactcccgagagag	414
Db	377	AAGCAGCCACG6CCCCAGATCATCTAATGTTACGTGCAABTCTTCAAGAGCTCCAGAGAG	436
OY	415	ccaacaatccagtgtaaacccccctgggcataccccctgtgaacaglaagagcctlgagagctc	474
Db	437	CCAACTATTCACCAAGCCAAACGTCGTGGGATCCACGTGACATACCAAGAGCTTAAAGGAGTT	496
OY	475	gtctaccctgtgtagggaggaagacgggttacccattcccttaagtcaatctgtgtacaagaatgc	534
Db	497	GCTACTCTGTGGGAGGAATAGGCTACCCCATTCCTCAACTCATCTGTGATCAAGAAATGAC	556
OY	535	cggcctctgaaggaggaanaaacccggtctcaattcaatcgtcctccaaagcctltggagtgc	594
Db	557	CGGCCCTTGCAAGAGAGAAACCGGGTTCATATCCAGTGTGCACAGACCGGTGGAGTCC	616
OY	595	agtgttgttiacacctctgcagaglatctctgaagcaacagcttgtttaaagaacaaagat	654
Db	617	AGTGGCTGTACACTTBAAGAGTGTCTTAAGTGTGCACGGGTATCTCAAGGAAGCAAGAT	676
OY	655	gccoaagtttctctgtgagctcaactacacccgctgcceaglyggaaacacatlgaagagatcc	714
Db	677	GCCCACTTTACTGTGAAGTACGTACCGGCCGCCCGCTGGAGAAACGGCATGTAAAGCAATCT	736
OY	715	agggaagtcacacgctccctglttcttaccggacaagaaaaagtgtgctltgaagtgtgagccc	774
Db	737	AAGGAGGTCACTGTCCCTGTCCTTATATCCCTGCGAGAAAAATGTGGGTGAGGTAGAACCC	796
OY	775	gtggaagatgcgtgaaggaagggagacccggtlgaatatcaagtggtlttgctltgatgcaacct	834
Db	797	GTGGGACTCTCTGAAGGAAGGGGACCATGTGAAGTCAAGGTGTGACTAAGGCAACCTT	856
OY	835	ccaaccaatctcagcatcagcaagaagaaccccaagcacaaccaaggacaggaagaagagaca	894
Db	857	CAACCCCACTTCACTATCAACAAACAAAGAACCCAGACACTAGSAGATGTGAAGAGGAGAC	916
OY	895	accaacgcgaacccgggtctcgtgtgtcgtgagacctgcgccgaagaaagaaacagtyggcgtat	954
Db	917	ACCCGATGAAAACGGGTCTCTCTCTCTTCCCTTGGACCCGCCAGAAAGCAACAGTGGGGCTAC	976
OY	955	gaatctcagggcccggaacttggaaccacatgatatcgtctgtctgtgtgaaccagaagaaata	1014
Db	977	CAATGTCAAGACCTTAGACTTBSAAACTACACGTTATGCTGTTCGAGTAACCCCTGGAGCTG	1036
OY	1015	ctgtgtgaactatgtgtctcgaagtcgcggtyagttcccgcaagcccccctlgaaagacaagaaggc	1074
Db	1037	CTGGHGAACATATGHTGTGTGATGTTCAAGTGTGATTCAAATCCCTCCCTGAAGTCCAGGAGST	1096
OY	1075	agcagcctccaccctcgacctgtgtgagcagaagatgagccaaggaactcgagttccaagtgcgt	1134
Db	1097	GATACCTTCACACACGACTGTGAAGGCGAAGATACCCAGACCTTGAAGTGTGCGCTG	1156
OY	1135	agagaagaagacagggccagagtgucctggaagaaggggccgtgtcttaagctgtcatagaactgaaa	1194
Db	1157	AGAGACAAGACAGGCTTCTTBGGAAGGGGCCCATCTTCAAGTTAAACAACGTGAAA	1216
OY	1195	cggagagcaagagcgycgtatcgtctgtgtgcgtctltgtgccacgatalaccgagatgaac	1254

OY	775	gttgaatgctggaaggaaggaaggaacgcgcgtggaataatcaagtggttttggcttgacacacct	834
Db	800	gttggggctgctgaaggaaggaagggatcatgtgacatccagtggtcttgacagatggcaacct	839
OY	835	ccaccaacacttcagcatcctgaagcaagcaaaccccccagccacagggagagcaaggaagaca	894
Db	860	caaccccacttctactatccaacaaagaaagaccgccactgsgggagatggaagaggaagc	919
OY	885	accacaagaacgcgggtctcgtggtgtgagactgtgccggaaggaacaagttggcgctat	954
Db	920	accatgtgaaataatggcgctcctgtcccttgagagcctcccaaaaagccactatccgggctcttac	979
OY	955	gaatgtcaagggccggagacttgagacacatgatcgtgtgtctggtgtagaccaaggaacta	1010
Db	980	cagttctcagaattcttgagactctggaatactacactgtgaactgaccccttgagacttg	1030
OY	1015	ctgtgtgaactatgtctctgacgtccagtgtagtcccgcaagcccttgagagacaaggaagc	1070
Db	1040	ctgggtgaactatgtgtgtcgtgaggtttcaagtgaaatccactgcctctgaatgcaggaaggt	1090
OY	1075	agcagcctcaccctgaacctgttgagcagaagtagccagaacctcgaattccagttgctg	1130
Db	1100	gagagccctcacgctcagactcctggaaagcagaaagatAACAGAGACCTTGAGTTGAGTGCCTG	1150
OY	1135	agaagaaggaaggaagggccaggtgtcttgaaaggggcgtgtgttaagttgactgacatga	1190
Db	1160	agagcaaaagaaagcagcagctgctcttggaagaaagggctccctctccaagctaaacacgtgaga	1210
OY	1195	cggagagcaaggaagcgcgtacacgtctgctgtgcgtctgtgcacagataaccggactaac	1250
Db	1220	cgggaagcagggggagcggatctctgtcatgsgcatctgtccccaagttccttgccgttgaa	1270
OY	1255	cgcaacaagctgtgtcaacgtgtgacatltttggcccccttgatgtagcatccaaggaag	1310
Db	1280	cgtaccacagcgtgcagcgtgtggcattttttgggtccccatgagatggcattaaaggaag	1330
OY	1315	aaggtgtgggtgaaggaagaaatagttttaaactgtgtcttgtagaagcgttaagggcacccc	1370
Db	1340	aaagttgtgggtgcgaagagaatgcagtgctgaattgtgtcttgtagagccttgaaagcatctt	1390
OY	1375	cgagcccaacatctctctggaacgttcaacgcgcagcagtgaaacaagaaccaagaatccacag	1430
Db	1400	cagccacacactctcctctggaatgtcaatggttcggcactgaaatggaacccaagatccacag	1450
OY	1435	cgagttcctggaacacccctggaatgtctcctgtgaccccggaagctgtgtggaagaagtgltgaa	1490
Db	1460	acagtagtagacacactctggaatgtccttctgtgacgcgaagagcttctggagagacaggtgcag	1510
OY	1495	tgcacggcctccaagagacgttggcaaaaaaacacagacatcctctctcgtggagcgtgtcaat	1550
Db	1520	tgttagcgtctcaactcctccttgsgctcaaaacacacacacatgttcttgaaagcttggtact	1570
OY	1555	ttaaacacacctcaacacagagactccaacaacaacaacctgtgctcagcaatctccactgtccagt	1610
Db	1580	tttaacacacctctatnaccnctgactccagcaaaacacactgsgcctcaacacacactccagtcagt	1630
OY	1615	cctataccagaagccaagaagacactccaagagaagaaagctgtgccggaagccggagaagccgg	1670
Db	1640	cctacacacacaagcacaacagacactccaacaaagaaaaagctgcccacacagagagcaaa	1690
OY	1675	ggcgtgagtcacagtcggtcgtgtatgtgtgacatccctgtgtcctggcggtgtgtgcgagctgc	1730
Db	1700	gggtgaggtcatcgttggtcgtgtatgtgtgtaactcttggtgtgtgtgcttgctgggtgctgct	1750
OY	1735	ctctatttctctataaagaaggcaagctgcgtgtgcagcgcgtccaaggaagcaaggaat	17930
Db	1760	ctctatttctctctaacaaagagaggaactgcctgcatgtgagagcctgggaaaaacaggaat	18180
RESULT	8		
GENECAML			
LOCUS	2106 bp	mrna	VMT
			06-FEB-1997

DEFINITION	G. gallus mRNA for HEMCAM, long cytoplasmic form.
ACCESSION	Y08855
VERSION	Y08855.1 GI:1621229
KEYWORDS	adhesion molecule; HEMCAM; immunoglobulin supergene family.
SOURCE	chicken.
ORGANISM	Gallus gallus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE	1 (bases 1 to 2106)
REFERENCE	Valinco, O., Dunon, D., Aissi, F., Dangy, J. P., McNagny, K. M. and Imhof, B. A.
TITLE	HEMCAm, an adhesion molecule expressed by c-kit+ hemopoietic progenitors
JOURNAL	J. Cell Biol. 135 (6 Pt 1), 1655-1668 (1996)
MEDLINE	97133433
REFERENCE	2 (bases 1 to 2106)
AUTHORS	Dunon, D.
TITLE	Direct Submission
JOURNAL	Submitted (15-OCT-1996) D. Dunon, Universite Pierre et Marie Curie, Unite CNRS 1135, Equipe Adhesion et Migration Cellulaire, 9 quat St Bernard, 75005 Paris, FRANCE
COMMENT	Related sequences: D38559 and D49849.
FEATURES	Location/Qualifiers
source	1..2106
	/organism="Gallus gallus"
	/strain="H.B19ov"
	/db_xref="taxon:9031"
	/dev_stage="E13"
	/tissue_type="thymus"
	201..305
	201..2081
	/note="long cytoplasmic form"
	/codon_start=1
	/product="HEMCAm"
	/protein_id="CAI70080.1"
	/db_xref="GI:1621230"
	/db_xref="SPTREMBL:Q98922"
	/translation="MAGGRRAPRCGCGMGCGCCPCLLLCCGAGRLVYMPAVL
	LEVIGESTARLECSFSIPGNASFTSIMFYVNRRLPISGEVNRHNASGVRIDETVESEE
	RLSAGEKALSLISKVTRODNARTEICGVNASQGVSGESRTELYTKIPAPEIPNSAA
	KIPGOSNDMLKIACTSENSPSPNITMYKNGEPLLOEDPKTLITLVRESNGLYIV
	VTLSFKVTRDRNSLPHCTHYHVALQCOMPRKDSPRVNVTFYFTEHVELRVATNAGI
	VKEDDYKLVYCDANGNPAPYPSFRRRLGDSWDMTSLATNDSVLMILHNYSKSSSL
	YRCOTLDLDDTQHGQVELVANTIEGVNAMESPLHEDSVRLSCVTAHSVPKLV
	QYRRARGRKAAEGQQLLTNLTFEFTSSNFSQVRAKSPVGLSDSKQYAAVAKKPRIV
	AISAPLYRVQDEYVINTLCKAIAFPQSPFHHMINTTHEYEMNGHMASNLVRYSHDL
	RAGMAGYRNALGVSEKHIDLDOKPESKIIIVALTIVCLVVAIVGSIYFLHKKGI
	KISGRSGKQDITKPEARKDKNVVEVMSDKISEAGLGLGANGEKSRPADQSEKVIDL
	RN"
BASE COUNT	306..2078
ORIGIN	472 a 596 c 679 g 359 t
	mat_peptide
	14.1%: Score 275.4; DB 4; Length 2106;
	Best Local Similarity 50.6%: Pred. No. 7.5e-50;
	Matches 986; Conservative 0; Mismatches 826; Indels 135; Gaps 8;
QY	18 ggtctgcgcctctctgtctgcgcgcgcgtgtctgtctgtctctgcgtctgcgcggtgtgtcccg 77
DB	245 GGGGGGGCGGCTGGGGCTGTGCGCCTTCTCTGCTGCTCTGCTGCTGCTGCGGAGCGCGG 304
QY	78 agagagctgaagcagcctgcgcctgaagctgtgtgaagtgtgaagtgtggcagacacaccccttc 137
DB	305 CAGGCTGGAGGCTCTACATGCGCAGCAGTCTCTGGAATGTGAGATCGGGACACAGCGAGCT 364
QY	138 gaagctgcgcctctcccaagtc---aaggaacaactgaagcaatgtcgcagtgttctgt 194
DB	365 GGAGTGCAGCTTCTTCATCCCTCGGGAATGCGCTCTTCACTCCATCCAGTGTGTTTATGCT 424
QY	195 ccaacaagagaagcagcagctcatcttcctgtgtgcgcgcagagccagagccagagcgaacc 254
DB	425 CAACCGCCGCGCCTCCATCGGAGTGAAGGCTGTGCGTATATACGCGCACGGGGTTCGCGAT 484

QY	225	tgaggagtaagagcagcggtcagcctccagagcaagagggtca---ctctgacctgac	311
Db	485	AGATGAACCCAGTACAGACGCGCGCTGTGCGTGGGGAGGACAAAGCCCTTCATCCAG	544
QY	312	tcgaagtcaccccccaagaagga-----gagatctctctgtgcagaggcaagcgcc	362
Db	545	CAAGGTGACCGGCAAGACAAACGCCAGGACCTTCATCTGCGACGTTGGGGGACAGCCA	604
QY	363	tcggtcccaaggagtaagccatccagctccgcgtctacaaagctccggaagagccaacat	422
Db	605	GGGCGTGGGGAGAGCCGCCACCGAGCTTACACTTACAAAGATCCCGGACCCCTTGAGT	664
QY	423	ccaagtcacacccccctgggcalccctgtgaacagttaaggagctgaggagtgctactctg	482
Db	665	CACACCCCAACTCGCGCGGCACTCCCTGCACAAAGCAATGATCTTGAAGATTCCTCAGTG	724
QY	483	tgtagggaggaagcggtaccccatctctcaagtcacatctgttaaaagaaagtcgcgctct	542
Db	725	CACGAGCGAAGAACAGCTTCCCGTCCCGCAATCAGTGGTATACAAAGCGGGAGCGCTT	784
QY	543	gaaggaggaagaagaccggygtccacatltcagtcgltcccaagactgtgagtcgagtgltc	602
Db	785	GCTGCAAGGAGGAGCAAGAACAAAGATCTGACACACGTTGTGTCGCGAGTCCGACGGCT	844
QY	603	gtacacctctgcaaggtatcttgaaggcacacgctggtttaaagaagacaaagatgcccagt	662
Db	845	GTACACGGTGTGTGACACTCTGTTTCAGCAAGGTGACACGGGAGACCGCAACTCCCTTT	904
QY	663	ttactgtgagcttcaactacacgctctgcacatgtggaaccacatgaagagttccagggaa	722
Db	905	CCACTGCAACCTGCACACTACTGTGCGAGGGACAGATGCGACCAAGGACTTCCGGGGT	964
QY	723	caccgtccctgtttcttctaccgcagacagaaaaagtgtgagctgtggaagtggagcc---cg	779
Db	965	GAACCTCACTGTGTTTACCCCAAGACAGACAGCTGAGACCTCGCGTGGCAACACCGCGG	1022
QY	780	aatgctgaaggaagaagggacggcggtgtgaaatcagagtglttgctgtgatagtgcaaccc	839
Db	1025	GATACTGGAAGGAAGGGATGATGTGAAGTTGTGCTGGATGCGATGGAAGGAAATCCGCGCG	1084
QY	840	aacactcagcatc-----aagcaagcagaacccagacacgaagcaggaagcaagagac	893
Db	1085	TGCTCTTTAGCTTCTTTCAGAGAGAGAGCTGGGGGACAGCTGGCAGGATATGATACCTCG	953
QY	894	aaccaaagcaacaggggtccctgtgtcgtgtgagcctctgcacggaaagaaacagtggtgc	953
Db	1145	GGATACCAACGATGAGTCTGATGCTGACAAATGTCAGCAAGAGCAGCAGCGGTCTGTA	1204
QY	954	tgaatgtcagggccttgagactctgagacacaaatgatatcgcgtgtgtgtgaacacagaact	1011
Db	1205	CAGGTGCCAGCACTGTGGACTTTGACGATTAAGCGCAGACGAAGGGGATGTGAGCTGTG	1266
QY	1014	actgtgtgaactatgtctctgtgcgtccagtgtagtcccgacgcccctcgagagacagaag	1077
Db	1265	GGTGAACATCAATTCGAAGGGGTCCAGGTGGAAGATGAGACCGTCTCTCAACCCCTTCATGAAG	1322
QY	1074	cagcagcctcaacccctgaactgtgagggcagaagtaagccaggacctcgaatltccagtgtct	1133
Db	1325	GGACAGCTGTGAGCTGTGAGTGCATCTGCCACACCCCGTCAAACTGGACTACCAATG---	1387
QY	1134	gaggaagagagcagggccaagtgctcgtggaagaggggctgtgtctcaatltgacgaactgt	1199
Db	1382	GAGGACCCCGAGGGGCGAGGAAGGTGGGAGGAAGCAACCACTTCCTTCAACCACTCCAC	1441
QY	1194	acggaagcaggaagcgagctatcgtctgctgtgagcgtctgtgcccagataccggagctga	1255
Db	1442	CTTGCAGACCTCCAGCAACTTCACTCTAAGATGGAAGGCTCGCAGTGTGCGGGGCTGGA	1501
QY	1254	ccgcaacacagcttgtaacgltgtgccaatltttggccccctctgtgatagtgcatlaaagaag	1311
Db	1502	GCAGGCGAAGCAGGTGGCGGTGCTGTGAAGGGGAAGCGCGCATGCTGGCATCAGCGC	1561

QY	1314	gaagtggtggttgaagaagaaatatygttgttaattcgtctgttttbtgaagcgtcaggacccc	1373
Db	1562	ccccgttagctacctggccgcagcgaagtgatcaactgacctgcatgaagccatccgtccctcc	1621
QY	1374	cgcggccaccacatctcccttggaagcgtcaacggccaaggcaagtgaagaccacgatcacca	1433
Db	1622	ccagccctctcttccactgtagaacatcaacgggagac---accacagatracatgacaacca	1678
QY	1434	gcgaqtlcttgagcacccatlgatatctctcgtgtgagcccgagatctgtttgagagcgtgttga	1493
Db	1679	gcacatgtgccagcaaacctgacggctgctgttgagaccagacacgactgctggccgacgagccat	1738
QY	1494	atgcacggcctcccaagaacctgtggcaaaaacaccagcatctctctcgtgagctgtgtaa	1553
Db	1739	gtgccgggtgtgtccaaacggctgggggtgcacgcgaagacacatccagctgctg-----	1790
QY	1554	ttaacacacctcaacacagactccaacaacacctgtcctcagacttccactgcgcag	1613
Db	1791	-----	1790
QY	1614	lccclatalccagagccacacagcaactcccaagagaaaaactgtccggagccggagagccg	1673
Db	1791	-----gattcaaaagccatccgacagcaa	1813
QY	1674	ggggcgtgtgcatcgtgtgctgtgtatgtgtgtcaatcctgtgtccgtggcgtgtgtcgtgt	1733
Db	1814	agggcatctatctgtgtggcatcatcgtctcatctcgtgtgtgtggccctgtccgtccat	1873
QY	1734	cctctatctctctataagaaggcaagctgcgcgtgcagcgctcagtgaggaagcaggagat	1793
Db	1874	catTTACTTCTTCGCACAGAAAGGCGAAGATCTCTCGGCCGACGGCAACAGGAAT	1933
QY	1794	caacgtgtcccccgtctcgtlaagagcgaaactgttagtltgaagtltaagtcagataagctcc	1853
Db	1934	cacaaagcccaagagcgagtaaacgaaagatgttagttaagttgaagtcagataaacttttc	1993
QY	1854	agaagaagatgggcctcctcgtcagggcagcgagtgacaagaagggtcccgaggaaccagag	1913
Db	1994	cgaaagcggcggtcctgcaggggtgcacagggcgagagaggtgcgcgcgtgacacagag	2053
QY	1914	agagaatacacatgactctgagggcata	1940
Db	2054	cgagaaatcatcatgactctgagaaacta	2080
RESULT	9		
LOCUS	CHKLCP	2584 bp	mRNA
DEFINITION	Chicken mRNA for I-glycerin precursor, complete cds.		VRT
ACCESSION	D49849		
VERSION	D49849.1	GI:1008910	
KEYWORDS	I-glycerin precursor; cell adhesion molecule; immunoglobulin superfamily;		
SOURCE	Gallus gallus (strain:Leghorn) adult gizzard cDNA to mRNA, clone_lid:lambda gt11.		
ORGANISM	Gallus gallus		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.		
AUTHORS	Taira,E.		
JOURNAL	Direct Submission		
TITLE	Submitted (24-MAR-1995) to the DDBJ/EMBL/GenBank databases. Elichih		
REFERENCE	Taira, Osaka University School of Medicine, Pharmacology; 2-2 Yamadaoka, Suita, Osaka 565, Japan		
AUTHORS	(E-mail:taira@pharm.med.osaka-u.ac.jp, Tel.:81-6-879-3521, Fax:81-6-879-3529)		
JOURNAL	2 (bases 1 to 2584)		
TITLE	Taira,E., Nagino,T., Taniura,H., Takaha,N., Kim,C.H., Kuo,C.H., Li,B.S., Higuchi,H. and Miki,N.		
REFERENCE	Expression and functional analysis of a novel isoform of glycerin,		
AUTHORS	an immunoglobulin superfamily cell adhesion molecule		
JOURNAL	J. Biol. Chem. 270 (48), 28681-28687 (1995)		
MEDLINE	96081930		

[illegible]

RESULT	10
LOCUS	HSWUC184
DEFINITION	H.sapiens MUC18 gene exons 6-13.
ACCESSION	X68267
VERSION	X68267.1 GI:602337
KEYWORDS	adhesion molecule; cell surface molecule; immunoglobulin superfamily; melanoma associated protein.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukariyota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS	Primates: Catarrhini; Homiidae; Homo.
TITLE	1 (bases 1 to 2015)
JOURNAL	Sers,C.
REFERENCE	Direct Submission
AUTHORS	Submitted (24-AUG-1992) C. Sers, Institut fuer Immunologie,
TITLE	Universitaet Muenchen, Goethestr 31, 8000 Muenchen 2, FRG
2 (bases 1 to 2015)	
Sers,C., Kirsch,K., Rothbacher,U., Riettmueller,G. and Johnson,J.P.	
Genomic organization of the melanoma-associated glycoprotein MUC18	
implications for the evolution of the immunoglobulin domains	
Proc. Natl. Acad. Sci. U.S.A. 90 (18), 8514-8518 (1993)	
93391384	
FEATURES	Location/Qualifiers
source	1..2015
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/cell_type="leucocyte"
	/clone_1ib="HL 106d (Clonetechn)"
	/clone="30B,17A,13"
	<1..147
Intron	/number=5
exon	148..327
	/number=6
Intron	328..447
	/number=6
exon	448..559
	/number=7
Intron	570..667
exon	/number=7
	668..830
Intron	/number=8
	831..925
exon	/number=8
	926..1044
Intron	/number=9
	1045..1146
exon	/number=9
	1147..1288
Intron	/number=10
	1289..1443

exon	/number=10 1444. 1565		
	/number=11		
intron	1566. 1676		
	/number=11		
exon	1677. 1818		
	/number=12		
intron	1819. 1915		
	/number=12		
exon	1916. 2011		
	/number=13		
intron	2012. >2015		
	/number=13		
BASE COUNT	424 a 592 c	606 g	393 t
ORIGIN			
Query Match	13.38;	Score 259.2;	DB 53; Length 2015;
Best Local Similarity	63.0%;	Pred. No. 2.5e-46;	
Matches 719; Conservative	0;	Mismatches 8;	Indels 415; Gaps 4;

OY	560	gggtccacattcaagtcgtgccagaagcttgaaftccagtggttgttgaaccacttgcagatga	613
Db	148	gggtccacattcaagtcgtgccagaagcttgaaftccagtggttgttgaaccacttgcagatga	207
OY	620	tctctgaagcacaagctggtttaaagaacaaagatgccagatttcactgtgaaccacat	679
Db	208	ttctgaagcacaagctggtttaaagaacaaagatgccagatttcactgtgaaccacat	267
OY	680	accggtctgcacgttgggaaccacaatgaaggatccagaagatcaacgctccgtttctc	739
Db	268	accggtctgcacgttgggaaccacaatgaaggatccagaagatcaacgctccgtttctc	327
OY	740	-----	739
Db	328	gtgaattactaacctgactcttctgcttacttgggcctgggctgcttgaaactctcttgcc	387
OY	740	-----	739
Db	388	ttcccccttatccaccgccagggttgactcatgggccaaagatcaacccctttgctctgaagc	447
OY	740	accgcagcagaaaagtgctgcttggaagtggaagccgttggaatgcttgaagaaaggaagcc	799
Db	448	accgcagcagaaaagtgctgcttggaagtggaagccgttggaatgcttgaagaaaggaagcc	507
OY	800	gcgttggaatcaagtgcttcttgcttgaatggaaccctccacaacatcgaacatcagaacag	859
Db	508	gcgttggaatcaagtgcttcttgcttgaatggaaccctccacaacatcgaacatcagaacag	567
OY	860	-----	859
Db	568	aggtgtgggggcttTgtGtGtgggtGcAGgtGggGatGCAAGCAGAGACCAAGTGGAGTG	627
OY	860	-----	881
Db	628	gcggaggtgacactTctCtCAcctccccaattctcccttgaAACcccaGACACAGGAGAGC	687
OY	882	agaagaaagaacaaaccaaagcaaacgaggtctcgtgtgtcgtgaagccttgcccgaaagaaca	941
Db	688	agaagaaagaacaaaccaaagcaaacgaggtctcgtgtgtcgtgaagccttgcccgaaagaaca	747
OY	942	cagtggtggtcctatgaatgtcaaggccttggaacttggaacccaatgatatcgcttgctgaagtga	1001
Db	748	cagtggtggtcctatgaatgtcaaggccttggaacccaatgatatcgcttgctgaagtga	807
OY	1002	accaacgaactactgttgaaact-----	1024
Db	808	accacagaaactactgttgaaact-----	867
OY	1025	-----	1026
Db	868	ggggcaattaggaacttgcccccaacccctgcgcctgggggagacatgattgctgtaccccccagat	927


```
misc_feature 132237..135290
/note="assembly_fragment"
misc_feature 135391..140229
/note="assembly_fragment"
misc_feature 140330..143896
/note="assembly_fragment"
misc_feature 143997..147572
/note="assembly_fragment"
misc_feature 147673..150446
/note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature 150547..153715
/note="assembly_fragment"
misc_feature 153816..157273
/note="assembly_fragment"
misc_feature 157374..159817
/note="assembly_fragment"
misc_feature 159918..161743
/note="assembly_fragment"
misc_feature 161844..163684
/note="assembly_fragment"
misc_feature 163785..165216
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature 165317..166657
/note="assembly_fragment"
misc_feature 166758..168307
/note="assembly_fragment"
misc_feature 168408..170030
/note="assembly_fragment"
misc_feature 170131..171725
/note="assembly_fragment"
misc_feature 171826..173038
/note="assembly_fragment"
BASE COUNT 45530 a 39606 c 41002 g 44200 t 2700 others
ORIGIN
```

```
Query Match 12.5% Score 243.2; DB 76; Length 173038;
Best Local Similarity 63.2% Pred. No. 4e-43;
Matches 722; Conservative 0; Mismatches 3; Indels 417; Gaps 6;
```

```
QY 560 ggggtccacatcagtcgtccagactgtagagtcgaagtggtttgtaacacctgacagagta 619
|||||
Db 33459 GGGTCCACATTCAGTGTCCAGACTGTGAGTGTGTTGTACACCTTGACAGAGTA 33400

QY 620 ttctgaaggacacagctggtttaaagaacaagaatgccagtttactgtgagctcaact 679
|||||
Db 33399 TTCTGAAGGACACAGCTGTTAAAGAACAAAGATGCCAGTTTACTGTGAGCTCACT 33340

QY 680 accgagctgccagtgagaaacacatgaaggagtcacaggaagtcacacgtccctgtttct 739
|||||
Db 33339 ACCGGGTGCCCACTGAGTGAACACATGAGAGATGCCAGGAGTACCGCTGTTTCT 33280

QY 740 ----- 739

Db 33279 GTGAGTACTAACTGACTCTCTGCTTACTGGGGCTGGGGCTCTGGAACCTCTGCTGGC 33220

QY 740 -----a 740
|
Db 33219 TCCCCCTTATCAGCCAGGTTGACTATGGGCCAAGAGCTACCCCTTTGCCCTATGCACA 33160

QY 741 cccgacagaaaaagtgtagctggaagtgagccgtgggaatgctgaaggaaggagacgg 800
|||||
Db 33159 CCCGACAGAAAAGTGTGGCTGGAATGAGCCCGTGGGAATGCTGAACGAAGCGGACCG 33100

QY 801 cgtggaaatcaggctgttgctgctgtagtgcaaaccttcacacacacttagatataagcaag- 859
|||||
Db 33099 CGTGAATAATCAGGTGTTGGCTGATGGAACCTCCACACACTTCAGATCAGCAAGCA 33040

QY 860 ----- 859

Db 33039 GGTGTGGGGCTTTTGTGCTGGTGCAGGTGGGATGCACAAAGACCAAGTGAGTGAG 32980

QY 860 -----agaaccacagcaccaaggagagca 882
```

```
|||||
Db 32979 CGGAGGTGCACTCTGTCTACACTTCCCAATTCCTCCCTAGAACCCCAAGACAGGAGGCA 32920

QY 883 gaggaaagacaacaacaacgaacaggggtccgtggtgagacctccggaaagaacac 942
|||||
Db 32919 GAGGAAGACAAACCAACGACGACGAGGGGTCTGTGTGTGAGGCTCCCGGAAGGAACAC 32860

QY 943 agtggagcatatgaatgtcacaggaccttgagacaccaatgaatcgtcgtgaatga 1002
|||||
Db 32859 AGTGGCGCTATGATATTCAGGGGCTGTGACTTGGACACCATGATTCGCTGATGATGA 32800

QY 1003 ccaaggaactactggtgaact----- 1024
|||||
Db 32799 CCACAGAGAACTACTGTGAACTGTGAGGGGTGGGACCCAGACAGAGGAGACAGGCTG 32740

QY 1025 -----atg 1027
|||||
Db 32739 GGGCAATAGGAACTGAGCCCAACCCCTGCTGGGTGACATGATTCGTAATCCCAAGATG 32680

QY 1028 tgtctgacgtccagtgatgctccgcaagccctgagagacagaaagcaagacctcacc 1087
|||||
Db 32679 TGTCTGACGTCCGAGTGAATCCCGACGCCCTGAGACAGAGGAGGAGGCTCA-CC 32621

QY 1088 tgaacctgtaggcagagaaagtagccaagacctgagttcagtggtcgaagaag----- 1141
|||||
Db 32620 TGACCTGTGA-GCAGAGAGTAGCCAGAGACTCGAGTCCAGAGCTGTGAGAAAGAGGTAC 32562

QY 1142 ----- 1141

Db 32561 CCAGGGGGGTCTGGGCGCAAGGGGCTGATGGGGTGGGCTGGCGGCTGACAAAGATG 32502

QY 1142 -----agaacagcagtgctggaag 1164
|||||
Db 32501 TGTCTTTCCAGACCTCTGACCATCTCTCATCCAGACAGGCGCAGTGTGGAAGG 32442

QY 1165 gggcctgtgctcagttgcatgacatgaaacggaggaagcggtctgctgctg 1224
|||||
Db 32441 GGGCTGTGCTTCACTGTGCAATGACCTGAACGGGAGGACAGAGGCGCTTTCCTGCTG 32382

QY 1225 ggcgtgtgcccagacatcacccggagactgaaccgcacacagctggtcagctggcatttt 1284
|||||
Db 32381 GCGTGTGGCCACATACCGGCTGAAACCGCACACAGCTGGTCAACGTGCCATTTT 32322

QY 1285 gg 1286
||
Db 32321 CG 32320
```

```
RESULT 12
GHEMCAAC 1988 bp mRNA VRT 06-FEB-1997
LOCUS G.gallus mRNA for HEMCAM, short cytoplasmic form.
DEFINITION Y08856
ACCESSION Y08856.1 GI:1621227
VERSION Y08856.1
KEYWORDS adhesion molecule; HEMCAM; immunoglobulin supergene family.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 1988)
Vainio,O., Dunon,D., Aissi,F., Dany,J.P., McNagny,K.M. and
Imhof,B.A.
HEMCAAC, an adhesion molecule expressed by c-kit+ hemopoietic
progenitors
JOURNAL J. Cell Biol. 135 (6 Pt 1), 1655-1668 (1996)
MEDLINE 97133433
REFERENCE 2 (bases 1 to 1988)
AUTHORS Dunon,D.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-1996) D. Dunon, Universite Pierre et Marie Curie,
Unite CNRS 1135, Equipe Adhesion et Migration Cellulaire, 9 quai St
Bernard, 75005 Paris, FRANCE
COMMENT Related sequence: D38559.
```


Db	1622	CCAGCCCTCCTTCCACTGGACATCAACGGGACGATGAGG---CTGGGGGGGTGCTGTG	1678
OY	1434	gcgagtcctcgagacacctgaatgtccctgtgaccccgagcgtgtgagacagtgltga	1493
Db	1679	GGCGGCTCTCGGGATGGGGGCTGTGGTGGCTGCTGAGCCGCCGTCTTCCTGTCCT	1738
OY	1494	atgcacgcctccacagacctgggcgaataaacacagcaltcccttcttgagctgttcaa	1553
Db	1739	ACAGACCCACACAGTACATGAGAACCCAGACATGGCCACCACACTCAAGCTGCTGTGAG	1798
OY	1554	tttaacacccctccacaccagactcccaacaacacactgtgctcagcacttccactgcag	1613
Db	1799	CCACGACCTGCTGCGGGCAGGAGCCA-----TGTGCCGGGTCTCAACGCGCTG	1847
OY	1614	tcctataccagagccaacagacactccacacagagagaacgtgcgagacgcgagagccg	1673
Db	1848	GGGGTCAGCGAAGACATCCACGTGCTGGATCAAAAGCCA-----TCGAGAGACAA	1900
OY	1674	ggggctgtgcatcgtgtgctgtatgtgtgtgcatctgtgtcctgtgcggtgtctggcgtgt	1733
Db	1901	AGGCAATCAATCATCGTGGCCATCATCTGTGCATCTGTGGTGGCGGTGCTGCTGCAT	1960
OY	1734	cctctattccctctataaaggcgcaagctgcgtgtgcagcgctcagggagagagagat	1793
Db	1961	CATTACTTCTCTGCACAGAAGGGCAAGATCTCTCTCGCGCCGACAGCGCAACAGGACAT	2020
OY	1794	cacg 1797	
Db	2021	AGCG 2024	
RESULT	14		
CHKSGP	CHKSGP	2365 bp	VRT 08-FEB-1999
LOCUS	Chicken mRNA for s-gicerin precursor, complete cds.		
DEFINITION	D38559		
ACCESSION	D38559.1	GI:1009246	
VERSION	s-gicerin precursor; cell adhesion molecule; extracellular matrix		
KEYWORDS	protein; neurite outgrowth factor-binding; laminin family; immunoglobulin superfamily; integral membrane glycoprotein.		
SOURCE	Gallus gallus (strain Leg Horn) adult gizzard (library: lambda		
ORGANISM	gtl1) cDNA to mRNA.		
REFERENCE	Gallus gallus		
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;		
TITLE	Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.		
JOURNAL	1 (bases 1 to 2365)		
REFERENCE	Taira, E.		
AUTHORS	Direct Submission		
TITLE	Submitted (20-Oct-1994) to the DDBJ/EMBL/Genbank databases. Elicht		
JOURNAL	Taira, Osaka University School of Medicine, Pharmacology; 2-2		
REFERENCE	Yamadaoka, Saita, Osaka 565, Japan		
AUTHORS	(E-mail:etaira@pharm.med.osaka-u.ac.jp, Tel:81-6-879-3521,		
TITLE	Fax:81-6-879-3529)		
JOURNAL	2 (bases 1 to 2365)		
REFERENCE	Taira, E., Takaha, N., Taniura, H., Kim, C.H. and Miki, N.		
AUTHORS	Molecular cloning and functional expression of gicerin, a novel		
TITLE	cell adhesion molecule that binds to neurite outgrowth factor		
JOURNAL	Neuron 12 (4), 861-872 (1994)		
REFERENCE	94213753		
MEDLINE	3 (sites)		
REFERENCE	Taira, E., Nagino, T., Taniura, H., Takaha, N., Kim, C.H., Kuo, C.H.,		
AUTHORS	Li, B.S., Hiyuchi, H. and Miki, N.		
TITLE	Expression and functional analysis of a novel isoform of gicerin,		
JOURNAL	an immunoglobulin superfamily cell adhesion molecule		
MEDLINE	J. Biol. Chem. 270 (48), 28681-28687 (1995)		
COMMENT	96081930		
FEATURES	On Oct 6, 1995 this sequence version replaced by:559700.		
SOURCE	updated (05-Oct-1995) by:Etichi Taira.		
	Location/Qualifiers		
	1..2365		
	/organism="Gallus gallus"		
	/strain="Leghorn"		
	/db_xref="taxon:9031"		

[illegible]

```

OY 683 ggcgtccacatggaaccacatgaagag tccagggaag tcaacctcctgtttcttacc 742
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 925 ggcgtcagggacagatggcgcacacacagacctctccggsgtgacacgtcattgtttcttacc 984
OY 743 cgaacgaagaagaatgtgtctgtgaagatgagc---cgtggaaatgtctgaagaagaaggacc 799
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 985 ccacagacagcagctggagcctggccgcacacacacccggcgatgtaggaagagggagatg 1044
OY 800 ggcgtggaatcgaagtg tttgtgctgtagtgaacccctccacacacttcaacatc-----a 853
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1045 atgtgaagattggtcttcgattgctgattggaaatccggccctgtcttttaccttttcacga 1104
OY 854 gcaacgaagaaccacacacacacacaggaagcagagaagaagaacaacacacacacaggaagc 913
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1105 gagagctggcgggacagctgcagcagatgatacctccgttggcgatgacaaacgattggagttcc 1164
OY 914 tgggtctgtgaagcctgcgcggaagaagaacagctggcgctatagtaatgtcaggagcctggact 973
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1165 tgatgttgacatgctacacacagacacacgctgtctgacgctgacgacacacgactgact 1224
OY 974 tgaacacacatactcgtcgtcgtgaagacacaggaactactgtgtgaactatgtgtcgt 1033
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1225 tggacacattgtgacgcacacacacagagggatgtgagctggtgactgactgacgaaagg 1284
OY 1034 acgtccgagtgagctcccgacaccccttgagagacaggaagcagacccctcaaccctgacct 1093
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1285 tccagctgaaatgagacgctgctcctcctcctcctcctcctcctcctcctcctcctcctcct 1344
OY 1094 gtagagcagaagatgaacaggaacacacacacacacacacacacacacacacacacacacag 1153
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1345 gacactgcccacacgcccgcgtcaagctgactgactgactgactgactgactgactgactgact 1401
OY 1154 tgcctgaaagaggagcctgtgtctcagcttgcacatgacacacacacacacacacacacacac 1213
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1402 agctggcgaagacgacacacacacacacacacacacacacacacacacacacacacacacac 1461
OY 1214 atcgtctgctgagcgtctgtgtccacacacacacacacacacacacacacacacacacacacac 1273
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1462 tcagctgacagatgaaagcttcgacagtggtccggcgctggagacagacacacacacacacac 1521
OY 1274 tggcacttttggccccctgtgacatgacacacacacacacacacacacacacacacacacac 1333
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1522 tggcgttgtagagggagacgctgctcagcttgcacatgacacacacacacacacacacacac 1581
OY 1334 atatagttgtgatactgtgtctgtgaagcgtcaagggcaccacacacacacacacacacacac 1393
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1582 acgaagtgatcaactgacacacacacacacacacacacacacacacacacacacacacacac 1641
OY 1394 acgtcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1453
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1642 ggcgtacacggcagc---acccacagatgacacacacacacacacacacacacacacacacac 1698
OY 1454 atgtcctcgtgaccccgagcgtgtgtgaagacagcgtgtgtgaatgacagcgtcccaagacc 1513
      |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 1699 cggctcgtgtgagcgcacacacacacacacacacacacacacacacacacacacacacacac 1758
OY 1514 tggg 1517
      |||||
Db 1759 tggg 1762

RESULT 15
AP001557 177595 bp DNA HTG 30-MAY-2000
LOCUS Homo sapiens chromosome 11 clone RP11-680A7 map 11q23, WORKING
DEFINITION DRAFT SEQUENCE, in unordered pieces.
ACCESSION AP001557
VERSION AP001557.2 GI:8117391
KEYWORDS HTG; HTGS; PHASE1; HTGS; DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-680A7.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 177595)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 177,595 genomic DNA of 11q23
Published Only in Database (2000) In press
2 (bases 1 to 177595)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission (29-MAR-2000) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gs.c.riken.go.jp, URL:http://hnp.gsc.riken.go.jp/
Tel:81-42-778-9923, Fax:81-42-778-9924)
On May 30, 2000 this sequence version replaced g1:7380892.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hnp.gsc.riken.go.jp/
Contact: hattori@gs.c.riken.go.jp
----- Project Information
Center project name: Humdrat11
Center clone name: RP11-680A7
----- Summary Statistics
Sequencing vector: PCR products, 100% of reads
Chemistry: Dye-terminator ET-amersham, 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 152340 bases at least Q40
Consensus quality: 165308 bases at least Q30
Consensus quality: 171688 bases at least Q20
Insert size: 174695; sum-of-ctrls
Quality coverage: 4.19x in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of
30 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

```

1 28368 contig of 28368 bp in length
28469 50818 contig of 22350 bp in length
50919 61024 contig of 10106 bp in length
61125 70311 contig of 8987 bp in length
70412 79377 contig of 8986 bp in length
79478 88609 contig of 9132 bp in length
88710 94861 contig of 6152 bp in length
94962 103809 contig of 8848 bp in length
103910 110603 contig of 6694 bp in length
110704 117602 contig of 6899 bp in length
117703 123711 contig of 6009 bp in length
123812 130010 contig of 6199 bp in length
130111 134042 contig of 3932 bp in length
134143 138065 contig of 3923 bp in length
138166 141127 contig of 2962 bp in length
141228 146317 contig of 5090 bp in length
146418 150591 contig of 4274 bp in length
150792 154193 contig of 3402 bp in length
154294 157006 contig of 2713 bp in length
157107 159710 contig of 2604 bp in length
159811 162554 contig of 2744 bp in length
162655 164588 contig of 1934 bp in length
164689 166160 contig of 1472 bp in length
167844 167844 contig of 1584 bp in length
167945 169717 contig of 1773 bp in length
171073 171073 contig of 1256 bp in length
171174 173241 contig of 2068 bp in length
173342 174526 contig of 1185 bp in length
174627 176098 contig of 1472 bp in length
176199 177595 contig of 1397 bp in length
Sequence updated (26-May-2000).

```

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved

FEATURES
Source Location/Qualifiers

1.177595
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/clone="RP11-680A7"
/map="11q23"
1.28368
misc_feature
/note="assembly_fragment"
28469..50818
misc_feature
/note="assembly_fragment"
50919..61024
misc_feature
/note="assembly_fragment"
61125..70311
misc_feature
/note="assembly_fragment"
70412..79377
misc_feature
/note="assembly_fragment"
79478..88609
misc_feature
/note="assembly_fragment"
88710..94861
misc_feature
/note="assembly_fragment clone_end:SP6 vector_side:left"
94962..103809
misc_feature
/note="assembly_fragment"
103910..110603
misc_feature
/note="assembly_fragment"
110704..117602
misc_feature
/note="assembly_fragment"
117703..123711
misc_feature
/note="assembly_fragment"
123812..130010
misc_feature
/note="assembly_fragment"
130111..134042
misc_feature
/note="assembly_fragment"
134143..138065
misc_feature
/note="assembly_fragment"
138166..141127
misc_feature
/note="assembly_fragment"
141228..146317
misc_feature
/note="assembly_fragment"
146418..150691
misc_feature
/note="assembly_fragment"
150792..154193
misc_feature
/note="assembly_fragment"
154294..157006
misc_feature
/note="assembly_fragment"
157107..159710
misc_feature
/note="assembly_fragment"
159811..162554
misc_feature
/note="assembly_fragment"
162655..164588
misc_feature
/note="assembly_fragment"
164689..166160
misc_feature
/note="assembly_fragment clone_end:17 vector_side:left"
166261..167844
misc_feature
/note="assembly_fragment"
167945..169717
misc_feature
/note="assembly_fragment"
169818..171073
misc_feature
/note="assembly_fragment"
171174..173241
misc_feature
/note="assembly_fragment"
173342..174526
misc_feature
/note="assembly_fragment"
174627..176098
misc_feature
/note="assembly_fragment"
176199..177595
misc_feature
/note="assembly_fragment"
BASE COUNT 44469 a 40734 c 42637 g 46855 t 2900 others
ORIGIN

Query Match 11.7%; Score 227.8; DB 76; Length 177595;
Best Local Similarity 67.9%; Pred. No. 8.9e-40;
Matches 468; Conservative 0; Mismatches 2; Indels 219; Gaps 2;

QY 560 gqgtccacattcagtcgtcccaagactgtgagtcgagtggtttgtaacctgcagata 619
|||||
Db 170194 GGGTCACATTTCAGTCGTCCAGACTGTGAGTCGAGTGTGGTACACCTTGACAGTA 170253
QY 620 ttctgaagcacagctgtgtlaagaagacaagatgccagtttactgtgagctcaact 679
|||||
Db 170254 TTCTGAAGCACAGCTGTTAAAGAGACAAGATGCCAGTTTACTGTGAGCTCAACT 170313
QY 680 accgctgcccagtggaagacacatgaagagagtcacaggaagtaacctcccttttct 739
Db 170314 ACCGCTGCCCCAGTGGGGAACACATGAAGAGTCCAGGAAAGTACCGTCCCTTTTCT 170373
QY 740 ----- 739
Db 170374 GTGAGTACTAAGCTGACTCTGCTTACTGCGGCTGGGCTGCTGAAGACTCTGCTGGC 170433
QY 740 ----- 739
Db 170434 TCCCCCTTATCCACGCCAGSGTTGACTCATGCGCCAAAGACTACCCCTTGGCTATGCA 170493
QY 740 -accgcacagaaaaagtcgtgagtcggaagtcgagccgtggaagtgcgaagaaaggaac 798
Db 170494 GACCCGACAGAAAAAGTGTGCTGGAAGTGAAGTGGAGCCCTGGGAATGCTGAAGAGAGGGAC 170553
QY 799 cgcgtgaaatacaagtggtttggtcgtatgcaaccctccacacactcagcatcaagaag 858
Db 170554 CGCGTGAATCAGGTGTTGGCTGTGATGCAACCTCCACACACTTCAGCATCAGCAAG 170613
QY 859 c----- 859
Db 170614 CAGGTGTGGGGCTTGTGTGCTGGGTGAGGTGGGATGCAAGACGACCCAGGTGAGT 170673
QY 860 ----- 880
Db 170674 GCGGGAAGTGCACCTGCTCACCCTCCCAATCTCCCTCGTAGAACCCAGACAGGAGG 170733
QY 881 cagaagaaagaaacaaacaaagacaaggggtcctgtgtgtgagcctgcgcgaagaac 940
Db 170734 CAGAGGAAGAGACACACCAACGACAGCGGGTCTGTGTGTGGAGCCTGCCGAAAGAAC 170793
QY 941 acagtggtcgtatgaatgtcaagggtcgtgacttgacacacatgatatcgtgtgagt 1000
Db 170794 ACAGTGGCGCTATGAATGTCAAGGGCTGAGACTTGACACACATGATATGCTGTCTAGTG 170853
QY 1001 aaccacaggaactactgtgtgaactatgtg 1029
Db 170854 AACCAACGAGAACTACTGTGTAACCTGTAG 170882

Search completed: March 26, 2001, 07:38:41
Job time: 6042 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2001, 06:35:54 ; Search time 105.4 Seconds
(without alignments)
6950.116 Million cell updates/sec

Title: US-09-653-961-1

Perfect score: 1950

Sequence: 1 atggggctcccaagctgtg.....tgaagcattagccccaat 1950

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_36:*

1: /cgn2_2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /cgn2_2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /cgn2_2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /cgn2_2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /cgn2_2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /cgn2_2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /cgn2_2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /cgn2_2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /cgn2_2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /cgn2_2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /cgn2_2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /cgn2_2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /cgn2_2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /cgn2_2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /cgn2_2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /cgn2_2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /cgn2_2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /cgn2_2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /cgn2_2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /cgn2_2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /cgn2_2/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1950	100.0	1950	20	220930	Human MUC18 CDNA s
2	1933.6	99.2	1960	20	220931	Human MUC18 CDNA s
3	1922.6	98.6	1962	20	220932	Human MUC18 CDNA s
4	44.6	2.3	681	15	Q72942	Monkey Erythropoie
5	44.6	2.3	1344	6	N50348	Monkey erythropoie
6	44.6	2.3	1344	19	V30955	Monkey erythropoie
7	44.6	2.3	5107	19	V21685	Vector plasmid CMV
8	42.8	2.2	10732	21	A10594	Gene encoding a su
9	42.6	2.2	2072	11	Q04529	Elastase gene, ps
10	42	2.2	5904	14	Q37813	Sequence encoding
11	42	2.2	5904	19	V42681	DNA encoding human
12	42	2.2	6575	16	Q84655	Human neuronal cal

13	42	2.2	6725	16	Q87834	Human neuronal cal
14	41.2	2.1	5059	10	X84332	Stealth virus nucl
15	38.8	2.0	925	10	N93196	Monkey recombinant
16	38.4	2.0	2148	21	Z49737	Mouse T110 membran
17	38.4	2.0	2249	9	N80921	Sequence encoding
18	37.8	1.9	1503	20	V72539	Therinus flavus amy
19	37.4	1.9	597	21	Z35142	Human immunoglobul
20	37.4	1.9	1477	21	Z98150	Human signal pepti
21	37.4	1.9	4496	17	T58553	Streptomyces prist
22	37.4	1.9	8051	18	T72685	Sugar biosynthesis
23	37.4	1.9	8160	20	X25774	S.erythraea erythr
24	37.2	1.9	1224	19	V60558	Gtld gene from Amy
25	37.2	1.9	1224	21	Z57050	A. orientalis glyc
26	37.2	1.9	1224	21	Z57051	A. orientalis glyc
27	37.2	1.9	1982	13	Q22748	Class II EPSPS gen
28	37.2	1.9	1982	18	T93788	Class II EPSP synt
29	37.2	1.9	1982	18	T77313	Class II EPSPS DNA
30	37.2	1.9	1982	18	V58009	Agrobacterium sp.
31	37	1.9	1079	21	Z99653	DNA coding for the
32	37	1.9	1995	12	O14184	N. clavipes draglin
33	37	1.9	1995	19	V23250	Nephila clavipes s
34	37	1.9	1995	21	Z38156	N. clavipes spider
35	36.8	1.9	180	16	Q97215	DNA used in the co
36	36.8	1.9	180	21	A03777	Bacteriophage fd p
37	36.8	1.9	1868	20	Z30409	Wheat M10 fungal r
38	36.8	1.9	2331	19	V07280	Rat semaphorin W e
39	36.8	1.9	4008	19	V07279	Rat semaphorin W e
40	36.6	1.9	861	19	V45116	A. fumigatus aller
41	36.6	1.9	1810	20	Z00802	Human secreted pro
42	36.6	1.9	2208	18	T62418	Z. japonica phosph
43	36.6	1.9	7435	21	Z93313	Sequence encoding
44	36.4	1.9	36519	19	V22141	Chimpanzee adenovi
45	36.2	1.9	1409	21	Z97090	Human secreted pro

ALIGNMENTS

RESULT	1	
ID	220930	standard; cDNA: 1950 BP.
AC	220930;	
DT	01-DEC-1999	(first entry)
DE	Human MUC18 cDNA sequence.	
KW	prostate cancer; melanoma; cell adhesion; glycoprotein metastasis; treatment; detection; diagnostic test; ds.	
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	CDS	1..1941
FT		/*tag= a
FT		/product= HumMUC18
FT		/note= "Human MUC18"
PN	W09945392-A1.	
PD	10-SEP-1999.	
PF	02-MAR-1999;	99WO-US04850.
PR	03-MAR-1998;	98US-0076664.
PA	(UYEM-) UNIV EMORY.	
PI	Wu G;	
WPI	1999-540899/45.	
P-PSDB	Y42404.	

Human neuronal cal
Stealth virus nucl
Monkey recombinant
Mouse T110 membran
Sequence encoding
Therinus flavus amy
Human immunoglobul
Human signal pepti
Streptomyces prist
Sugar biosynthesis
S.erythraea erythr
Gtld gene from Amy
A. orientalis glyc
A. orientalis glyc
Class II EPSPS gen
Class II EPSP synt
Class II EPSPS DNA
Agrobacterium sp.
DNA coding for the
N. clavipes draglin
Nephila clavipes s
N. clavipes spider
DNA used in the co
Bacteriophage fd p
Wheat M10 fungal r
Rat semaphorin W e
Rat semaphorin W e
A. fumigatus aller
Human secreted pro
Z. japonica phosph
Sequence encoding
Chimpanzee adenovi
Human secreted pro

QY 1741 ttccctataagaaggcaagctgcccgtgcaagcgctcaaggaggaagcagatcacgctg 1800
 CC |||||||
 CC ttccctataagaaggcaagctgcccgtgcaagcgctcaaggaggaagcagatcacgctg 1800
 Db 1741 ttccctataagaaggcaagctgcccgtgcaagcgctcaaggaggaagcagatcacgctg 1800
 QY 1801 cccccgtctgtaagaagcgaacttctagttgaattgaattcaagtcacgaatagctcccaagaag 1860
 CC |||||||
 CC cccccgtctgtaagaagcgaacttctagttgaattgaattcaagtcacgaatagctcccaagaag 1860
 Db 1801 cccccgtctgtaagaagcgaacttctagttgaattgaattcaagtcacgaatagctcccaagaag 1860
 QY 1861 atggagctctgcaagcgaacgctgcaagaagcgctcccggaagaccaggaagagaa 1920
 CC |||||||
 CC atggagctctgcaagcgaacgctgcaagaagcgctcccggaagaccaggaagagaa 1920
 Db 1861 atggagctctgcaagcgaacgctgcaagaagcgctcccggaagaccaggaagagaa 1920
 QY 1921 tacatcgatctgagcattagcccgaaat 1950
 CC |||||||
 CC tacatcgatctgagcattagcccgaaat 1950
 Db 1921 tacatcgatctgagcattagcccgaaat 1950

RESULT 2
 ID 220931 standard; cDNA; 1960 BP.
 XX
 AC 220931;
 XX
 DT 01-DEC-1999 (first entry)
 XX
 DE Human MUC18 cDNA sequence, previously published (Johnson 1989).
 XX
 KM Prostate cancer; melanoma; cell adhesion; glycoprotein;
 XX metastasis; treatment; detection; diagnostic test; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 8..1948
 FT /tag= a
 FT /product= HumMUC18
 FT /note= "human MUC18"
 FT
 XX
 PN WO945392-A1.
 XX
 PD 10-SEP-1999.
 XX
 PF 02-MAR-1999; 99WO-US04850.
 XX
 PR 03-MAR-1998; 98US-0076664.
 XX
 PA (UYEM-) UNIV EMORY.
 XX
 PI Wu G;
 XX
 DR WPI; 1999-540899/45.
 XX
 DR P-PSDB; Y42405.
 XX
 PT Detection of metastatic prostate cancer, by detection of MUC18
 XX expression in prostate cancer cells -
 XX
 PS Disclosure: Page 8; 80pp; English.
 XX
 CC This is the previously published nucleotide sequence of the Human MUC18
 CC cell adhesion glycoprotein, which is expressed on the surface of
 CC melanoma cells, and can be used as a marker for prostate cancer.
 CC This nucleotide sequence is virtually identical to the humMUC18 sequence
 CC set forth in the invention, except for seven amino acid residues (which
 CC may be due to allelic differences). However the amino acid sequence of
 CC Johnson's sequence was 43 residues shorter than the sequence proposed
 CC by the inventors.
 CC The presence of this glycoprotein has been correlated with the ability of
 CC melanomas to metastasize. MUC18 is also associated with normal vascular
 CC tissue, and on the smooth muscle of venules, and it expresses
 CC sporadically on capillary epithelium.
 CC The method can be used as a diagnostic test for prostate cancer which
 CC has a relatively high potential for metastasis or which has metastasized.
 CC The physician can then choose the appropriate surgical, chemotherapeutic

CC or radiation treatment regimens. In addition, antibody specific to MUC18
 CC can be used to prevent metastasis of Prostate Cancer Cells.
 CC
 XX Sequence 1960 BP; 457 A; 557 C; 597 G; 349 T; 0 other;
 SQ

Query Match 99.2%; Score 1933.6; DB 20; Length 1960;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 1939; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 atggagctcccaagcgtctgcaagcgaacttctagttgaattgaattcaagtcacgaatagctcccaagaag 60
 Db 8 atggagctcccaagcgtctgcaagcgaacttctagttgaattgaattcaagtcacgaatagctcccaagaag 67
 QY 61 gtgcgagggtgtgcccggaagagctgtagcagcctgcgctgaagctgtgtgaggtgagagtg 120
 Db 68 gtgcgagggtgtgcccggaagagctgtagcagcctgcgctgaagctgtgtgaggtgagagtg 127
 QY 121 ggcagcacagccctctgaaatgtagcctctcccaagcccaaggaaccccaagcattgc 180
 Db 128 ggcagcacagccctctgaaatgtagcctctcccaagcccaaggaaccccaagcattgc 187
 QY 181 gactggtttctgtccacaagaagaaagcgagcgtcattcttcgtgtgagccagggccag 240
 Db 188 gactggtttctgtccacaagaagaaagcgagcgtcattcttcgtgtgagccagggccag 247
 QY 241 ggcagagcgaaacctgaggagctacgagcagcgctcagcctccagagcaagaggctact 300
 Db 248 ggcagagcgaaacctgaggagctacgagcagcgctcagcctccagagcaagaggctact 307
 QY 301 ctggccctgactcaagtaaccccccaagaagagcagcattcttctgtccagggcaagcgc 360
 Db 308 ctggccctgactcaagtaaccccccaagaagagcagcattcttctgtccagggcaagcgc 367
 QY 361 cctcggtcccaaggagtaacgcagcagctccgcgtctcaacaagctcccgaggagccaaac 420
 Db 368 cctcggtcccaaggagtaacgcagcagctccgcgtctcaacaagctcccgaggagccaaac 427
 QY 421 atccaagttcaacccccctgggcatccctgtgaacaagtaagaagcctgagaagctgcgtacc 480
 Db 428 atccaagttcaacccccctgggcatccctgtgaacaagtaagaagcctgagaagctgcgtacc 487
 QY 481 tgttgaaggaggaacgggtgtaaccccaattcccaagtaactctgtgaacaagaataggccgct 540
 Db 488 tgttgaaggaggaacgggtgtaaccccaattcccaagtaactctgtgaacaagaataggccgct 547
 QY 541 ctgaagagaggaagaacccggtccacattcagtcgtcccaagctgtgaagtcgaatggt 600
 Db 548 ctgaagagaggaagaacccggtccacattcagtcgtcccaagctgtgaagtcgaatggt 607
 QY 601 ttgtacaaccttgcaagatattctgaaggcacaagctgtgttaagaagaacaagaatgcccag 660
 Db 608 ttgtacaaccttgcaagatattctgaaggcacaagctgtgttaagaagaacaagaatgcccag 667
 QY 661 tttaacgttagctcaactacccggtctgccagtgggaacacacacatgaagagtgctcaggaa 720
 Db 668 tttaacgttagctcaactacccggtctgccagtgggaacacacacatgaagagtgctcaggaa 727
 QY 721 gtccagctccctgttttctacccgacagaaaagtgtgtgtgaagtgaagtcggcgaggaa 780
 Db 728 gtccagctccctgttttctacccgacagaaaagtgtgtgtgaagtgaagtcggcgaggaa 787
 QY 781 atgtgtgaaggaaggagccgctgtggaatacagtggtttgtgctatgcaacccctcacca 840
 Db 788 atgtgtgaaggaaggagccgctgtggaatacagtggtttgtgctatgcaacccctcacca 847
 QY 841 cacttcagcatcagcaagcagaaccccaagcaggaaggaaggaaggaagcaaaccaac 900
 Db 848 cacttcagcatcagcaagcagaaccccaagcaggaaggaaggaaggaagcaaaccaac 907
 QY 901 gacaacggggtcctgtgtgtgagcctgtccggaaagaaacacagtggtgctatgaaagt 960
 Db 908 gacaacggggtcctgtgtgtgagcctgtccggaaagaaacacagtggtgctatgaaagt 967

[illegible]

Db	1270	caagctcgttcaagcctgtcccaatttttggccccctttgatttgatctgaagatccaaggagaaggaagt	1325
Qy	1321	tgggtgaaagagaataatagttgttgaatactgtctctgttgaagcgttcaaggacccccggccc	1380
Db	1330	tgggtgaaagagaataatagttgttgaatactgtctctgttgaagcgttcaaggacccccggccc	1389
Qy	1381	accatctctctggaacgttcaacgagcagcgaaagtgaacaagaccaaagatccacagcgaatc	1440
Db	1390	accatctctctggaacgttcaacgagcagcgaaagtgaacaagaccaaagatccacagcgaatc	1449
Qy	1441	ctggaacaccttgaatagtctctgttgacccccggagctggttggaaacaggtgttgaatgcag	1500
Db	1450	cttggacaccttgaatagtctctgttgacccccggagctggttggaaacaggtgttgaatgcag	1509
Qy	1501	gctctcaagaccttggcaaaaacacagatccctctctctgaagctgttgaatgaattaac	1560
Db	1510	gctctcaagaccttggcaaaaacacagatccctctctctgaagctgttgaatgaattaac	1569
Qy	1561	acccttcaacacagacttcaacaacaacacatgctgcttcagacttcaacttccagttccat	1620
Db	1570	acccttcaacacagacttcaacaacaacacatgctgcttcagacttcaacttccagttccat	1629
Qy	1621	accagaagccaagaagaacacttcccaagaagaaaagctgcgcgaagccggaagccggggcgtg	1680
Db	1630	accagaagccaagaagaacacttcccaagaagaaaagctgcgcgaagccggaagccggggcgtg	1689
Qy	1681	gtcacgtgtgctctgtatttgttgcatctccgtgctccgtgacgtgtctgagcgttctctat	1740
Db	1690	gtcacgtgtgctctgtatttgttgcatctccgtgctccgtgacgtgtctgagcgttctctat	1749
Qy	1741	ttcctctaaagaagaagcgaagctgtccgttgcgaagcgtctcaaggaaacgaagatacagctg	1800
Db	1750	ttcctctaaagaagaagcgaagctgtccgttgcgaagcgtctcaaggaaacgaagatacagctg	1809
Qy	1801	cccccgctctcgtaaagaagcgaacttgttagttgaagttcaagtaagctcccaagaag	1860
Db	1810	cccccgctctcgtaaagaagcgaacttgttagttgaagttcaagtaagctcccaagaag	1869
Qy	1861	atcgagctctctcaagcagcagcgtgtgaacaagagctcccgagaccaggaagaga	1920
Db	1870	atcgagctctctcaagcagcagcgtgtgaacaagagctcccgagaccaggaagaga	1929
Qy	1921	tacatcgatctg-aggcattagccccga	1948
Db	1930	tacatcgatctgaggcattagccccga	1958
RESULT 4			
O72942/c			
ID	O72942	standard; DNA; 681 BP.	
XX			
AC	O72942;		
XX			
DT	30-AUG-1995	(first entry)	
XX			
DE	Monkey Erythropoietin cDNA.		
XX			
KM	Primer: PCR; amplify: human; mouse; erythropoietin; BP; mutain;		
KW	replacement; stimulation; erythrocyte; proliferation; cell culture;		
XX	hypertension; seizure; headache; nausea; ss.		
OS	Macaca mulatta.		
XX			
FT	Key	Location/Qualifiers	
FT	CDS	72..569	
FT		/tag= a	
FT		/product= Erythropoietin	
XX	W09424160-A.		
PN			
PD	27-OCT-1994.		
XX			
PP	21-APR-1994;	94WO-US04361.	

FT XX /*tag= c
PN XX A0688723-B.
PD XX 19-FEB-1998.
XX XX
XX XX
PF 02-DEC-1997; 97AU-0046867.
XX XX
PR 02-DEC-1997; 97AU-0046867.
XX XX
PA (KIRI) KIRIN AMGEN INC.
XX XX
PI Lin F;
DR WPI: 1998-261957/24.
DR P-PSDB; W58399.
XX XX
PT Recombinant human erythropoietin - potentially useful for diagnosis
PT and treatment of blood disorders
XX XX
PS Example 3; Page 34-36; 100pp; English.
XX XX
CC The present sequence encodes monkey erythropoietin (EPO), from an
CC example of the present invention. The present invention describes
CC recombinant human EPO which causes bone marrow cells to increase
CC production of reticulocytes or red blood cells, where the polypeptide
CC is the product of expression in CHO (Chinese hamster ovary) cells of
CC an exogenous DNA sequence encoding human EPO. EPO is potentially
CC useful in the diagnosis and treatment of blood disorders
CC characterised by low or defective red blood cell production.
XX XX
SQ Sequence 1344 BP; 280 A; 408 C; 400 G; 256 T; 0 other;

Query Match 2.3%; Score 44.6; DB 19; Length 1344;
Best Local Similarity 50.7%; Pred. No. 0.078;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 1059 tgagagacaagaagcagcaccctcagctgtagagagagagtagcagagacct 1118
DB 617 TGATGCTTCGGAGTGGAGCAGCCGCGCATCTGGAGAGGAGATGCTTCTGGGCTC 558
QY 1119 cgaagtcagtgctgagagagagagagcagcagtgctggaaggagggcctgcttca 1178
DB 557 CCAGCGCCCGAAGCAGATGATGATGTCGCAAGGCCACTGATGCTTATCCATGTGCA 498
QY 1179 gttgcatgactgaagcagggagggagggcgctatcgctgctgctgctgctccag 1238
DB 497 GCTGCAGGGGCTCGAAAGGCTGGGAAGTTGGCCACACAGCGCCCTGGCCGACAGAC 438
QY 1239 cataccggaactgaacgcagacacagctgctc 1269
DB 437 CTTCTGAGAGCAGCGGCCAGGCCCTGCCAGAC 407

RESULT 7
V21685/c
-ID V21685 standard; DNA: 5107 BP.
XX
XX V21685;
XX
XX 17-AUG-1998 (first entry)
XX
XX Vector plasmid CMV-KM-cMEPO.
XX
XX Polynucleotide delivery; plasmid CMV-KM-cMEPO; vector;
XX gene therapy; vaccine; polyclonal agent; cynomolgus monkey;
XX erythropoietin; EPO; anaemia; ss.
XX
XX Chimeric - Macaca cynomolgus.
XX Chimeric - Human cytomegalovirus.
XX Chimeric - Bos taurus.
XX

PN W09806437-A2.
XX 19-FEB-1998.
XX
XX
XX 13-AUG-1997; 97WO-US14465.
XX
XX
XX 13-AUG-1996; 96US-0023867.
XX
XX
XX (CHIR) CHIRON CORP.
XX
XX Cohen F, Dubois-Stringfellow N, Dworki V, Innis MA;
XX Murphy JE, Tetsuo U, Zukermann R;
XX WPI: 1998-159296/14.
XX
XX Polyclonic agents based on alpha-amino acids, able to complex
XX with nucleic acid - to facilitate its entry into cell, condense it
XX and protect it against serum degradation, particularly for use in
XX gene therapy
XX
XX Example 7; Page 81-83; 100pp; English.
XX
XX This polynucleotide sequence comprises plasmid CMV-KM-cMEPO, a
XX vector in which cynomolgus monkey erythropoietin (Epo) cDNA
XX (ATCC 67545) is inserted into the XbaI and BamHI sites of the
XX mammalian expression vector CMVkm2 (see V21684) under control of
XX the human cytomegalovirus immediate early gene promoter/enhancer and
XX the polyA site of the bovine growth hormone gene. Hematocrit
XX levels increased in SCID mice injected with the vector 1.m.
XX PCMV-KM-cMEPO is an example of a vector that can be used in novel
XX compositions and methods for improved polynucleotide delivery into
XX cells. In these methods, polyclonic agents are used to increase
XX the frequency of uptake of a nucleic acid (see also V21683-86) into
XX a cell. The polyclonic agent can condense with the nucleic acid
XX and inhibit serum and/or nuclease degradation of the nucleic acid.
XX The nucleic acid can be a vector, may express a therapeutic protein
XX or a vaccinating viral or cancer antigen, or is itself therapeutic
XX (antisense or ribozyme). The methods and compositions can be used
XX in the gene therapy of many diseases.
XX
SQ Sequence 5107 BP; 1145 A; 1463 C; 1335 G; 1164 T; 0 other;

Query Match 2.3%; Score 44.6; DB 19; Length 5107;
Best Local Similarity 50.7%; Pred. No. 0.13;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 1059 tgagagacaagaagcagcaccctcagctgtagagagagtagcagagacct 1118
DB 2187 TGATGCTTCGGAGTGGAGCAGCCGCGCATCTGGAGAGGAGATGCTTCTGGGCTC 2128
QY 1119 cgaagtcagtgctgagagagagagagcagcagtgctggaaggagggcctgcttca 1178
DB 2127 CCAGCGCCCGAAGCAGATGATGCTCGAAGGCCACTGATGCTTATCCATGTGCA 2068
QY 1179 gttgcatgactgaagcagggagggagggcgctatcgctgctgctgctgctccag 1238
DB 2067 GCTGCAGGGGCTCGAAAGGCTGGGAAGTTGGCCACACAGCGCCCTGCCAGACAG 2008
QY 1239 cataccggaactgaacgcagacacagctgctc 1269
DB 2007 CTTCTGAGAGCAGCGGCCAGGCCCTGCCAGAC 1977

RESULT 8
A10594
-ID A10594 standard; DNA: 10732 BP.
XX
XX A10594;
XX
XX 29-JUN-2000 (first entry)
XX
XX Gene encoding a subunit of cellulose synthase.
XX
XX

XX —

XX —

[illegible]

FT	/tag=	a
FT	/note=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT		
FT	misc_difference 3610	b
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3615	c
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3630	d
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3631	e
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3638	f
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3641	g
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3642	h
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3657	i
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3659	j
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3681	k
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3698	l
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3703	m
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3708	n
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		
FT	misc_difference 3713	o
FT	/tag=	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	/note=	
FT		

FT	misc_difference	3720	/*tag=	p	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3732	/*tag=	q	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3733	/*tag=	r	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3740	/*tag=	s	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3747	/*tag=	t	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3750	/*tag=	u	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3751	/*tag=	v	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3766	/*tag=	w	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3778	/*tag=	x	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3792	/*tag=	y	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3799	/*tag=	z	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3800	/*tag=	aa	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3812	/*tag=	ab	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3813	/*tag=	ac	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT			/note=		
FT	misc_difference	3818	/*tag=	ad	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT					

F7	/note=
FT	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
ET	
misc_difference	3819
FT	/tag=
FT	/note=
FT	a*
FT	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	
misc_difference	3824
FT	/tag=
FT	/note=
FT	a*
FT	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	
misc_difference	3827
FT	/tag=
FT	/note=
FT	a*
FT	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	
misc_difference	3841
FT	/tag=
FT	/note=
FT	a*
FT	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	
misc_difference	3853
FT	/tag=
FT	/note=
FT	a*
FT	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	
misc_difference	3857
FT	/tag=
FT	/note=
FT	a*
FT	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	
misc_difference	3860
FT	/tag=
FT	/note=
FT	a*
FT	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
FT	
misc_difference	3863
FT	/tag=
FT	a]
FT	"this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"

Query Match	2.18;	Score 41.2;	DB 20;	Length 5059;
Best Local Similarity	17.0%;	Pred. No. 0.92;		
Matches 36;	Conservative 123;	Mismatches 149;	Indels 2;	Gaps 1;
QY 144	cgagctctccagctcccaaggaacctccagccatgctgcagctgtttctgcgcacaga	203		
Db 145	csammrnscanradatcgrvrsbydstlyngncdcsttlttrsaargnyrssym	204		
QY 204	gaagcgagcctatcttccgtgtgctgcgaagggccagggccagagcgaaccttgaggaga	263		
Db 205	tmaanaactvrnctmmayccordnrgthlurdrldtgnratamarycarmmrnsandc	264		
QY 264	cgagcgagctcagagctccagagacagagggactctc-tggcctgactaagtaacc	321		
Db 265	anasbabydctccarmandamagncldvranctcdsnasayvrnctscancctthmn	324		
QY 322	ccccagaagagcgagctctctgtgtgccagggcgaagcgccctcgltcccaagaagtaacc	381		
Db 325	sytslmaadramasaarslmaatlryclmthraycndhtnctwthmammmndcncy	384		
QY 382	atccagctcgcgctctacaaagctccggagaggaacaaatcccaagtcgaacccctgggc	441		
Db 385	rshvrscgntcdctsnmmncmtncrtlarvrsayasnrlrrthmmngcadnscthyg	444		
QY 442	atccctctggaacgaaggaagcctcgagag	471		
Db 445	hscmhchansmasaawngatlntrshs	474		

Search completed: March 26, 2001, 07:37:16
Job time: 3682 sec

```

XX AC N93196;
XX DT 28-JUN-1990 (first entry)
XX DE Monkey recombinant kallikrein gene.
XX KW Monkey; kallikrein gene; vasodilator; male infertility;
XX OS Monkey.
XX FH Key Location/Qualifiers
XX FT misc_feature 39..
XX FT /tag= a
XX FT /note="start of monkey kallikrein sequence"
XX FT 78..848
XX FT /tag= b
XX FT /note="mature protein"
XX EP297913-A.
XX PN 04-JAN-1989.
XX PD 30-JUN-1988; 88EP-0306039.
XX PF 30-JUN-1987; 87US-0068594.
XX PR 30-JUN-1987; 87US-0068594.
XX PA (AMGE-) AMGEN INC.
XX PI LIn FK, Lu HS;
XX DR WPI: 1989-009139/02.
XX DR P-PSDB; P93722.
XX PT New recombinant kallikrein polypeptide(s) and encoding DNA -
XX PT useful as vasodilators and for treating male infertility.
XX PS Table IV; page 11; 43pp; English.
XX CC The monkey kallikrein gene is 93% homologous to that of the
XX CC human genomic kallikrein gene. See also P92314, N93194,
XX CC N93195, N93193, N93197, N93196, P93719, P93720, P93721, and
XX CC P93722.
SQ Sequence 925 BP; 209 A; 281 C; 243 G; 192 T; 0 other;

```

```

Query Match 2.0%; Score 38.8; DB 10; Length 925;
Best Local Similarity 46.9%; Pred. No. 1.9;
Matches 121; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 692 gtgggaaccacatgaagagtcacggggaagtcacccgtgttctacccgacagaa 751
   ||||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 217 gtgggggcatccctgtgcatcccccagtggtgtcaccagctgcatcagcgaca 276
QY 752 aagtggtgctggaagtcgacccgtggaatgctgaaggaagccgctggaatca 811
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 277 attaccagctgtgctgtgctgcacaaactgtttgatgaagaaacacagccca 336
QY 812 ggtgttgcgtatgcagcaacctccacacacatlcagcatcagaagaaacccaga 871
   | | | | | | | | | | | | | | | | | | | | | | | | |
Db 337 ttcatgtcagtgagagcttccacacacctgtgttcaacatgagcctcctgaagaacaca 396
QY 872 ccagggaagcagaggaagacaacccaagcaacggttcctggtgtgagcctgccc 931
   || | | | | | | | | | | | | | | | | | | | | | | | |
Db 397 ccgcgcaagcagatgattacagcagacactcattgtctccgcctgaagcagcctgcg 456
QY 932 ggaaggaacacagtcggc 949
   || | | | | | |
Db 457 agatcacagacgctgtgc 474

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 26, 2001, 05:41:24 ; Search time 1105.8 Seconds
(without alignments)
12357.208 Million cell updates/sec

Title: US-09-653-961-1
Perfect score: 1950
Sequence: 1 atggggcttccagctgtf.....tgaggcatagcccggaat 1950

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estp11:*
70: em_estp12:*
71: em_estp13:*
72: em_estp14:*
73: em_estp15:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: em_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp19:*
 121: em_estp20:*
 122: em_estp21:*
 123: em_estp22:*
 124: em_estp23:*
 125: em_estp24:*
 126: em_estp25:*
 127: em_estp26:*
 128: em_estp27:*
 129: em_estp28:*
 130: em_estp29:*
 131: em_estp30:*
 132: em_estp31:*
 133: em_estp32:*
 134: em_estp33:*
 135: em_estp34:*
 136: em_estp35:*
 137: em_estp36:*
 138: em_estp37:*
 139: em_estp38:*
 140: em_estp39:*
 141: em_estp40:*
 142: em_estp41:*
 143: em_estp42:*
 144: em_estp43:*
 145: em_estp44:*
 146: em_estp45:*
 147: em_estp46:*
 148: em_estp47:*
 149: em_estp48:*
 150: em_estp49:*
 151: em_estp50:*
 152: em_estp51:*
 153: em_estp52:*
 154: em_estp53:*
 155: em_estp54:*
 156: em_estp55:*
 157: em_estp56:*
 158: em_estp57:*
 159: em_estp58:*
 160: em_estp59:*
 161: em_estp60:*
 162: em_estp61:*
 163: em_estp62:*
 164: em_estp63:*
 165: em_estp64:*
 166: em_estp65:*
 167: em_estp66:*
 168: em_estp67:*
 169: em_estp68:*
 170: em_estp69:*
 171: em_estp70:*
 172: em_estp71:*
 173: em_estp72:*
 174: em_estp73:*
 175: em_estp74:*
 176: em_estp75:*
 177: em_estp76:*
 178: em_estp77:*
 179: em_estp78:*
 180: em_estp79:*
 181: em_estp80:*
 182: em_estp81:*
 183: em_estp82:*
 184: em_estp83:*
 185: em_estp84:*
 186: em_estp85:*
 187: em_estp86:*
 188: em_estp87:*
 189: em_estp88:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	651	33.4	690	135	BE786708	BE786708 601475064
2	624.6	32.0	1034	110	BE619435	BE619435 601473262
3	609.4	31.3	975	137	BE879940	BE879940 601491164
4	603.8	31.0	637	105	BE257461	BE257461 601111413
5	596.2	30.6	669	28	AL042548	AL042548 DKEP4341
6	551.2	28.3	679	106	BE275425	BE275425 601121423
7	532.6	27.3	553	136	BE871209	BE871209 601448791
8	513.8	26.3	594	136	BE813148	BE813148 RCI-BN000
9	466.4	23.9	715	135	BE786166	BE786166 601474277
10	461.2	23.7	729	14	AA983120	AA983120 ua35d07.r
11	442.6	22.7	564	105	BE234610	BE234610 141772 MA
12	435.2	22.3	602	105	BE226518	BE226518 ia21907.Y
13	422.6	21.7	642	93	AM611108	AM611108 un77a02.Y
14	410	21.0	642	93	AM611108	M78206 EST01799 Su
15	399	20.5	423	3	AA191148	AA191148 zp86c01.r
16	382	19.6	527	2	AA088962	AA088962 mm49e11.r
17	371.8	19.1	382	5	AA325490	AA325490 EST28484
18	368.2	18.9	563	8	AA522327	AA522327 v145d03.r
19	337.6	17.3	528	21	AI528914	AI528914 ud42h04.Y
20	319.2	16.4	403	5	AA323568	AA323568 EST26450
21	317.6	16.3	468	19	AI325641	AI325641 mm49e11.Y
22	311.8	16.0	403	108	BE484000	BE484000 170309 BA
23	303.8	15.6	442	2	AA085691	AA085691 zn63d06.r
24	302.8	15.5	442	87	AW231129	AW231129 uc39h07.Y
25	300	15.4	433	108	BE482927	BE482927 168887 BA
26	292.4	15.0	315	5	AA322791	AA322791 EST25565
27	289.6	14.9	502	18	AI283948	AI283948 q171g04.x
28	285.8	14.7	402	108	BE487749	BE487749 176829 BA
29	279.2	14.3	388	136	BE863695	BE863695 UT-M-BHO-
30	272.8	14.0	353	23	AI664124	AI664124 ue82b06.r
31	271.4	13.9	353	108	BE485906	BE485906 173198 BA
32	268	13.7	532	136	BE851455	BE851455 uw96d03.Y
33	267.8	13.7	313	90	AM392972	AM392972 CMI-STO27
34	264.8	13.6	268	26	AI902799	AI902799 RC-BT015-
35	264.2	13.5	269	26	AI902803	AI902803 RC-BT015-
36	261.2	13.4	306	133	BE006161	BE006161 RCO-BN012
37	255.6	13.1	272	6	AA349269	AA349269 EST55868
38	236.2	12.1	262	97	AA997154	AA997154 RC2-BN004
39	225	11.5	466	14	AA989790	AA989790 uc79b03.Y
40	220.2	11.3	465	142	N39748	N39748 vx29h08.r1
41	216.8	11.1	340	105	BE226755	BE226755 us71p08.Y
42	216.8	11.1	450	142	N31224	N31224 yx3d05.r1
43	213	10.9	512	105	BE257665	BE257665 601113092
44	212	10.9	254	3	AA187703	AA187703 zp72p12.r
45	208.4	10.7	237	136	BE838197	BE838197 CMI-FN010

ALIGNMENTS

RESULT 1
 BE786708 BE786708 690 bp mRNA
 LOCUS 601475064F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3678246 5',
 DEFINITION mRNA sequence.
 ACCESSION BE786708
 VERSION BE786708.1 GI:10207906
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 690)
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DRP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM628 row: o column: 15
High quality sequence stop: 673.
Location/Qualifiers
1..690
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3878246"
/clone_1ib="NIH-MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPO6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 168 a 175 c 228 g 119 t
ORIGIN

Query Match 33.4%; Score 651; DB 135; Length 690;
Best Local Similarity 98.1%; Pred. No. 6.2e-153;
Matches 680; Conservative 0; Mismatches 10; Indels 3; Gaps 2;

QY 681 ccgagtcgcccgaatggaacacataaagaatccagggaagtcaccgctcctgttttcta 740
DB 1 cgggctgcccgaatggaacacacataaagaatccagggaagtcaccgctcctgttttcta 59
QY 741 ccggaacgaacaaagtgtgctggaagtcgaagccgtggaatgctggaaggaaggaagc 800
DB 60 cccgacagacaaagtgtgctggaagtcgaagccgtggaatgctggaaggaaggaagc 119
QY 801 cgtggaatcaggtgttggctgcatggaacccctccacacacatcagcatcagaagca 860
DB 120 cctgcaaatcaggtgttggctgcatggaacccctccacacacatcagcatcagaagca 179
QY 861 gaacccacacacaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 920
DB 180 gaacccacacacaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 239
QY 921 gaagcctgcccgaaggaacacacaggtggtcatatgctcagggccctggaactggaac 980
DB 240 ggaagcctgcccgaaggaacacacaggtggtcatatgctcagggccctggaactggaac 299
QY 981 catgatalcgtctgagtgaaacacaggaactactgtgtgtaactatgtgtgtaactc 1040
DB 300 catgatalcgtctgagtgaaacacaggaactactgtgtgtaactatgtgtgtaactc 359
QY 1041 agtgaagtcggaagccctggaaggaaggaaggaaggaaggaaggaaggaaggaag 1100
DB 360 attgagtcggaagccctggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 419
QY 1101 aagaagtagcaggaagcctggaagtcagtgctggaaggaaggaaggaaggaaggaag 1160
DB 420 aagaagtagcaggaagcctggaagtcagtgctggaaggaaggaaggaaggaaggaag 479
QY 1161 aaggagggcgtgtgctcaggtgcatgacacccaaggaaggaaggaaggaaggaagga 1220

DB 480 AAGGGGGCGCTGCTGCTTCATGATGACCTGAACGGGAGGAGGAGGGCGCTATCGCTG 539
QY 1221 cgtgagctgtgtccagcatatccaggactgaagcagacagctgtgtaagtgccat 1280
DB 540 CGTGCGCTGTGTGCTGAGATACCCGG--CTGAACCCGACACAGCTGTCAACGTGCCAT 597
QY 1281 ttgtggccccccttgatgcatcgaaggaaggaaggaaggaaggaaggaaggaagga 1340
DB 598 TTTGGCCCCCTTGATGAGGATTCAGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 657
QY 1341 gtgaatcgtctgtgtgaagcgtcagggcacc 1373
DB 658 GTTGAATCTGTCTGTGAAGCGCTCGGGACCC 690

RESULT 2
BE619435 1034 bp mRNA EST 24-AUG-2000
LOCUS 601473262F1 NIH-MGC_68 Homo sapiens cDNA clone IMAGE:3876255 5',
DEFINITION mRNA sequence.
ACCESSION BE619435
VERSION BE619435.1 GI:9890373
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1034)
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DRP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM623 row: 1 column: 16
High quality sequence stop: 678.
Location/Qualifiers
1..1034
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3876255"
/clone_1ib="NIH-MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPO6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 296 a 276 c 308 g 153 t 1 others
ORIGIN

Query Match 32.0%; Score 624.6; DB 110; Length 1034;
Best Local Similarity 97.7%; Pred. No. 2.9e-146;
Matches 708; Conservative 0; Mismatches 9; Indels 8; Gaps 7;

QY 359 ggcctggtccaggaagtagcgaatccgcgtclacaaagtcctggaaggaagca 418
DB 1 GCCCTCGGTCCAGAGATACCGATCCAGCTCCGCTTACAAAGCTCCGGAGAGACCAA 60
QY 419 acatccaggtcaaccccttgatcctctgtgaacgttaaggaaggaaggaaggaagga 478
DB 61 ACATCCAGGTCAACCCCTTGATCCCTGTGTAACAGTAAGAGCCTGAGAGGCTGCTA 120
QY 479 cctgtgaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 538

Db 121 CCTGTGATGAGGAGGAACGGGTACCCATTCTCAAGTCATCTGTGTACAAAGAAATGCGCGC 180

QY 539 cctctgaagaaggagaagaacccgggtccacattcagtcgtcccaagactgtgagtcgagtg 598

Db 181 CTCTGAAGGAGGAGGAACCGGGTCCACATTCAGTCGTCGCCAGACTGTGAGTGCAGTG 240

QY 599 gttctacacactctgacagatattcttaagcacagctgtgttaagaagaacaagatgcc 658

Db 241 GTTTGTACACTCTTGACAGATATTCTGAAGGCACAGCTGGTTAAGAAGCAAGAATGCC 300

QY 659 agtttactgtgagctcaactacacggtctgcccagctggaacacatgaagaagtcacagg 718

Db 301 AGTTTACTGTGAGCTCAACTACCGGCTGCCAGTGGGAACCATGAGAGAGTCCAGGG 360

QY 719 aagtcacgtccctgttttctacccgacagaagaagtgtgctgtg - aagltgagccctgt 777

Db 361 AACTCACCCGCTCTGTTTCTTACCCACAGAAAAGTGTGCTGGCAAGTGGACCCCGTG 420

QY 778 ggaatctggaagaaggagccgctggaatcaggctgtttgtgtatgtgcaacccctca 837

Db 421 GGAATGCTGAAGACCGGGACCCGCTGGAATCAGCTGTTGGCTGATGCAACCCCTCA 480

QY 838 ccacactcagcaatcgcaagcaagaaccccaagacacaggaagcagaagaagacacc 897

Db 481 CCACACTTCAGCATCAGCAAGCAAGACCCACGACCCAGGAGCAGAGAGAGACAA-C 539

QY 898 aagcaaaacgggtgtctgtgtgtgagcgtcccggaagaagacagatgtggcgtatga 957

Db 540 AACGACACACGGGGTCTCGGTGCTGAGCCTG -CCGGAAGGACACAGTGGCCCTATGAA 598

QY 958 tgtcagggccctgtg -acttggacacatgataltcgttc -tgaatgaacacacagaa -ct 1013

Db 599 TGTCAGAGGCTGGGACTGTGGACACCATGATTCGCTGTGATGACACACGAACT 658

QY 1014 actgtgtgaactatgtgtcgtgacgtccgagtgagtc -gcagcccttgagagaagaag 1072

Db 659 ACTGTGAATATGTGTGACGTCCGAGTCCGGACCCCTGGAGAGACCGGAG 718

QY 1073 gcagc 1077

Db 719 GCAGC 723

RESULT 3

BE879940 975 bp mRNA EST 27-SEP-2000

LOCUS BE879940

DEFINITION 601491164F1 NIH_MGC_69 Homo sapiens cDNA clone IMAGE:3893607 5', mRNA sequence.

ACCESSION BE879940

VERSION BE879940.1 GI:10328716

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 975)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTB/DRP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCM668 row: o column: 16
High quality sequence stop: 645.

FEATURES

source location/Qualifiers

1..975

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3893607"
/clone_lib="NIH_MGC_69"
/tissue_type="large cell carcinoma, undifferentiated"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-SPORT6; Site: 1; Note: Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."

BASE COUNT 278 a 224 c 317 g 156 t

ORIGIN

Query Match 31.3%; Score 609.4; DB 137; Length 975;
Best Local Similarity 95.8%; Pred. No. 1.9e-142;
Matches 723; Conservative 0; Mismatches 21; Indels 11; Gaps 9;

QY 549 ggaagaagaaccggttcacatcagtcgtcccaagactgtgagtcgagtcgtgtacac 608

Db 1 GGAGAAAGAACGGGGTCCACATTCAGTGTCTCCACAGCTGTGAGTGCAGTGGTTGTACAC 60

QY 609 ctctcagaatctctgaagcagacagctgtgttaagaagaacaagatgccagtttactg 668

Db 61 CTTCAGAGATATTCTGAAGGCACAGCTGTAAAGAAACAAATGCCCCAGTTTACTG 120

QY 669 tgagctcaactacacggctgcgcacgttgggaaccacatgaagaagtcacagggaagtcacgt 728

Db 121 TGAGCTCAACTACCGGCTGTGCCAATGSSAACCATATGAGAGTCCA -GGAAGTCACTGT 179

QY 729 cccgttttctaccccgacagaagaatgtgtgtgagtgagccgttggaaatgtcgtgaa 788

Db 180 CCTGTTTCTTACCCGACAGAAAAAGTGTGCTGAAGTGAAGCCCTGGGAATGCTGAA 239

QY 789 ggaagaagacggcgtgtggaatcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 848

Db 240 GGAAGGGGACCGCGTGAAATCAGGTGTTTGCTGTATGAGCAACCTCCACACACTTCAG 299

QY 849 catcagaacagcaaaccccaagacaggaaggac - ggaagaagacaaacacagcaacag 907

Db 300 CATCAGCAAGCAAGAACCCACAGCACAGGAGGACAGAGAAAGAACAAACACACAGCAACG 359

QY 908 gggctctgtgtgtgagcctgtcccgaaagaaacagtggtggtatgtatgtatgtatgtatgt 967

Db 360 GGGTCTGTGCTGTGAGGCTGCCCGAAGAAACAGTGGCGCTATGAAATGATGAGGGCC 419

QY 968 tggacttgagacacccatgtatcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1027

Db 420 TGGACTTGGACACCAWGTATGCTGTGTGAGTGAACCAAGAACTACTGTGAACTATG 479

QY 1028 tgtctgacgtccgaatgtagtcocagcagcccttgagaagacaggaagcagcagctcacc 1087

Db 480 TGTCTGACGTCCGACTGATGATCCCGCAG -CCCTGAGAGACAGAAAGGACAGCTCAACC 538

QY 1088 tgaactgtgaagcagagatgtagcagacgtcgtatccagtggtcgtgaagaagaagacag 1147

Db 539 TGAACCTGTGAGGACAGAGATGAGCCAGGACCTCGAG -TCCACTGCTGTGAGAAAGACAG 597

QY 1148 gccacgtgtcgtgaagaagggt 1207

Db 598 GCCAGGTGTGAAAGGGGG -CTGTGCTTCAAGTTGCAATGA -CTGAAGGGAGGACAGGAG 655

QY 1208 gggatcatgcgtgcgt 1265

Db 656 GCGG -TATCGGTAGGGTGTGGTTGTGCCAAGATTAACCGGGGCGGAAACAAAGACT 714

QY 1266 gttcaagcttgacatltttggccccccttgagtg 1300

Db 715 GGTCAACGTTGGC -ATATTGGCCCCCTTGGGTTG 747

RESULT 4

BE257461

LOCUS BE257461 637 bp mRNA EST 13-JUL-2000
 DEFINITION 60111413F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3352034 5',
 mRNA sequence.
 ACCESSION BE257461
 VERSION BE257461.1 GI:9127936
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 637)
 NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
 Plate: L10M150 row: j column: 03
 High quality sequence stop: 601.
 Location/Qualifiers
 1..637
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3352034"
 /clone_1ib="NIH_MGC_16"
 /tissue_type="retinoblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: eye; Vector: pOT7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH-MGC library."
 BASE COUNT 125 a 195 c 199 g 118 t
 ORIGIN
 Query Match 31.0%; Score 603.8; DB 105; Length 637;
 Best Local Similarity 99.4%; Pred. No. 4,4e-141;
 Matches 627; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 361 cctcgtlccccaagagatccagatccagctccgcgtctctacaaagctccgagagcccaac 420
 |||||||
 Db 361 CCTCGGTCCCAAGAGATACCGCATCCAGCTCCGGGTCTACAAAGCTCCGAGAGCCCAAC 420
 QY 421 atccaggtcaacccctggatccctgtgacacagtaagagcctcgagaggttcgtacc 480
 |||||||
 Db 421 ATCCAGGTCAACCCCTGGGCATCCCTGTGAACAGTAAGACCTCAGAGAGTCCGTACC 480
 QY 481 tcttgtagaggaagacgggagaccatctcctcaagatcgtgtacagaatggccgacct 540
 |||||||
 Db 481 TGTGTAGGAGAGACCGGATACCCCAATTCCTCAAGTCATTTGTTACAGATGGCGCCT 540
 QY 541 ctgaagaggaagaaagaccgggt-ccacatcagtcgtctccagactgtgagtcagttg 599
 |||||||
 Db 541 CTGAAGGAGAGAAAGAACCGGGTCCACATTCAGTGTGCCAGAGTGTGAGTGCAGTGG 600
 QY 600 ttgtacaccttgacagatattctgaagca 630
 |||||
 Db 601 TTGTG-ACACCTTGACAGATATTCTGAAGCA 630
 RESULT 5
 AL042548 669 bp mRNA EST 29-FEB-2000
 LOCUS AL042548
 DEFINITION DKFZp43412321_r1 434 (synonym: htes3) Homo sapiens cDNA clone
 DKFZp43412321 5', mRNA sequence.
 ACCESSION AL042548
 VERSION AL042548.1 GI:5422003
 KEYWORDS EST.
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 669)
 AUTHORS Blum, H., Bauersachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
 TITLE EST (Blum, et al.)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Blum H
 MIPS
 Am Klopfersplitz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by LMU (Ludwig Maximilians University,
 Munich/Germany) within the cDNA sequencing consortium of the German
 Genome Project.
 No SI sequence available.
 This clone (DKFZp43412321) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Location/Qualifiers
 1..669
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp43412321"
 /clone_1ib="434 (synonym: htes3)"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
 BASE COUNT 159 a 197 c 182 g 130 t 1 others
 ORIGIN
 Query Match 30.6%; Score 596.2; DB 28; Length 669;
 Best Local Similarity 99.0%; Pred. No. 3.6e-139;
 Matches 620; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1322 gggtgaagaagaatagtgtgtaatctgtctgtgaagcgtcagagaccccgagccca 1381
 |||||||
 Db 1 GGGTGAAGAAGAAATATGATGTTGAATCTGTGTGAAGCGTCAAGGACCCCGGCCCA 60
 QY 1382 ccattcctgtaagcgaagcagcgaagtgaaacaagacaaagatccacagcgagtc 1441

```
|||||
Db 61 CCATCTCTGGAACGTCAAGCGACGCAAGTAAACAGCCAAAGTCCACACGCGATCC 120
OY 1442 tggagcaaccctgaaatgctcctgagcccgagagctgcttgagagacagtgcttgaatgcacg 1501
Db 121 TGAGCAACCTGGAATGTCCTCTGTGACCCCGAGCTGTTGGAGACAGTGTGAATGACAGG 180
OY 1502 cctccaaagcagcttgycgaaaaacacagacatcctctccttgagcgcgtgcaatttaacca 1561
Db 181 CCTCAACAGCACTGGGCAAAACACAGCATCTCTTCTGAGCGTGTCAATTTAACCA 240
OY 1562 cctctaaacagagctccaaacaaacacacagcctgagcctcagcacttcagctccatcctcata 1621
Db 241 CCTCAACAGCACTGCAACACACACAGCTGCGCTCAGACCTTCCAGTCCAGTCTCATTA 300
OY 1622 ccaagaaaccaagcagcactccacagagagaaagctgcgagagccgagagagccgagag-cgtg 1680
Db 301 CCAAGAGCAACAGCACTCCACAGAGAGAAAGCTGCCGGAGCCGGAGAGCCGGGGCCGTG 360
OY 1681 gtacatcgtgagctgagctgtgagcaccctgagctcctgagcgtgctgagcgtgtcctctat 1740
Db 361 GTCATGCTGGCTGTGATGTGTGTCATCTGTGCTGCTGCGGCTGTGGCGCTGTCTCTAT 420
OY 1741 ttcctctataagaagagcagagctgcccgtgcaagcgtcagaggaagcagagatcacgctg 1800
Db 421 TTCCTCTATTAAGAGAGGCAAGCTGCGGTGCAAGGCTCAAGGAGAGAGATCACGCTG 480
OY 1801 ccccgctctcgttaagaagcagagctgtgagttgaagttaaagtcagataagctcccaagaag 1860
Db 481 CCCCCGCTCTGTAAGAGAGCAAGCTTGTAGTTGAAGTTAAGTCAGATTAAGCTCCCAAGAG 540
OY 1861 atggagcctcctgcaagcagcagcagctgagcaagaagagc-tccgagagaccagagagagaa 1919
Db 541 ATGGAGCTCTCTGCAAGGACACAGCGGTGACAAAGAGGGCTTCCGGAGACAGAGAGAGAA 600
OY 1920 atacatcgatcgtgagcattagcccc 1945
Db 601 ATATCATGATCTGAGGCAATTAGCCC 626

RESULT 6
LOCUS BE275425 679 bp mRNA EST 13-JUL-2000
DEFINITION 601121423F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:2988868 5',
ACCESSION BE275425 mRNA sequence.
VERSION BE275425.1 GI:9150379
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 679)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LCM78 row: n column: 05
High quality sequence stop: 588.
Location/Qualifiers
1..679
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2988868"
/clone_lib="NIH_MGC_20"
```

```
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: skin; Vector: pOT87; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'-
adaptor: GGCAGGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 159 a 196 c 197 g 127 t

ORIGIN

Query Match 28.3%; Score 551.2; DB 106; Length 679;
Best Local Similarity 96.1%; Pred. No. 7.2e-128;
Matches 640; Conservative 0; Mismatches 18; Indels 8; Gaps 7;

```
OY 143 ggggctctccagtcacagcaactcagcatalgtc-gactggtttc-tgtccaaa 200
Db 1 GGGGCTCTCCAGTCCAGTCCAGCAAGCAACTCAGCAATGCTGATGCTTCTTCTGTCACAA 60
OY 201 ggaagaagcgaagcctcatctccgtgtgctgagcgaagggccca-gagcgaacctggg 259
Db 61 GGAGAAAGGGAGAGCTCATCTTCCGTGTGCGCCAGGGCCAGGCGCATGAGCGAACTTGGG 120
OY 260 agtaagagcagcagctcagcctcagagcagaggggtactctgacctgaactca 319
Db 121 AGTAGAGACAGCGGTGACGCTCCAGAGACAGAGGGGCTACTCTGCGCTGACTCAATCA 180
OY 320 ccccccagaagcagcactctctctgtgacagggcaagcgccctcgtgtcccaagagtaac 379
Db 181 CCCCCAAGAGAGAGAGCGCATCTTCTTGTGTCAGGGCAAGCGCCCTGCTCCAGAGAGTACC 240
OY 380 gcatcagctcccggtctcaaaaagctccggagggccaaaatccag-gtcaaacccctg 438
Db 241 GCATTCACGCTCCGGCTCTTACAAAGCTCCGGAGGAGCCAAACATCTCACTGTCACACCCCTG 300
OY 439 -ggatccctgtgaacagtaagagcctgagagctgactacacctgtgtaggaagaa 497
Db 301 TGCGATCCCTGTGAACAGTAAAGACCTGAGAGAGGTCCCTACTCTGTAGGAGAGAACGG 360
OY 498 gtaccacatctcgaatcattgtgtacaagaatgagccgtcttgaagagagaa 557
Db 361 GTACCCCATTCCTCAAGTCAATCTGTACAAAGATGGCGGCTCTGAAGAGAGAGAA 420
OY 558 ccgggtccacattcagtcgtgtcccaagcgtgtgagtgagtggtgtgtacacctgcagag 617
Db 421 CCGGGTCCACATTCAGTGTCTCCAGACTGTGAGTTCAGTGTGTGACACTTGGCAGAG 480
OY 618 tatctgaagcagacagctggttgaagaagacaagaatgcccagtttactgtgagctcaa 677
Db 481 TATTTGTGAAGGACACAGCTGTGTCAAAAGAACCAAGATGCCAGTTTACTGTAGCTCAA 540
OY 678 ctacc-ggctgcccagtgaggaacacatgaagaggtcccaaggaatcaccgtccctgtt 736
Db 541 CTACCTGATGATCCAGTGTGAACCATGAGAGAGTCAAGAGAGTCAACGATACCTGTTA 600
OY 737 tctaccgcaagaaaagtgtgtgtgtgagagtgagagcccggtgagaaatgctgaagaagg 796
Db 601 TCAACCCCAAGACAAAGTGTGGCTTGAAGT--AGCCGTGGGCAATGTGAAGACGGA 658
OY 797 accgag 802
Db 659 CGCGTG 664

RESULT 7
LOCUS BE871209 553 bp mRNA EST 27-SEP-2000
DEFINITION 60144791F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852671 5',
ACCESSION BE871209 mRNA sequence.
```

VERSION BE871209.1 GI:10319985
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 553)
 AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLC562 row: e column: 24
 High quality sequence stop: 551.

FEATURES
 source
 1..553
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3852671"
 /clone_lib="NIH-MGC-65"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: NotI;
 Site: 2: SalI; Cloned unidirectional. Primer: Oligo dT.
 Average insert size 1.8 Kb. Library constructed by Life
 Technologies."

BASE COUNT 146 a 142 c 174 g 91 t
 ORIGIN

Query Match 27.3%; Score 532.6; DB 136; Length 553;
 Best Local Similarity 98.9%; Pred. No. 3.2e-123;
 Matches 547; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

```

QY 643 gaagaaagaatgcccagtttactgtgagctgaactacggctccagtggaaccac 702
    |||||||
DB 1 GAAGACAAGATGCCCCAGTTTACTGTGAGCTCACTACCGGCTGCCAGTGGAAAC 60
QY 703 atgaagagatccagggaatccgcctcctgtttctaccacgaagaagtgtgctg 762
    |||||||
DB 61 ATGAAGAGATCCAGCAAGTCAACGCTCCTGTTTCTACCCGACGAAAAAAGTGGCTG 120
QY 763 gaagtggagcccggtggaaatgctgaagaaaggagccgctggaatcaggtgttgc 822
    |||||||
DB 121 GAAGTGGAGCCCGTGGGATGCTGAAGGAAGGGACCGGTGGAATCAGTGTGGCTT 180
QY 823 gatgtgaacccctccacacattcagcatcagcaagcagaacccacagcaggaag 882
    |||||||
DB 181 GATGGCAACCTCCACACACTTCAGCATCAAGCAAGAAACCCACGACGAGGAGCA 240
QY 883 gaggaagaaacaacaagaagggtgctggtgctggaagcctgcccgaagaacac 942
    |||||||
DB 241 GAGGAAGAAACAACCAACGACGAGGAGGCTGCTGGTGGAGCCCGGGAAGGAAAC 300
QY 943 agtgggagctatgaatgtcgaaggcctggaactggaacacatatactgctgagtgaa 1002
    |||||||
DB 301 AGTGGGCGCTATGAATGTCAAGGGCTGGAGCTTGACACCATATATTCCTGCTGAGTAA 360
QY 1003 ccacaggaactctgtgaactatgttctgaagtcgagtgagtgagtcggcagccctgag 1062
    |||||||
DB 361 CCACAGGAACACTGCTGTAATGTGTGACGTCGACAGTCAAGTCCGCGAGCCCTGAG 420
QY 1063 agacaggaagcagagcctcacccctgagcctgtgagcagagagatgagcagaactcag 1122
    |||||||
DB 421 ACACAGGAAGGACGACGCTCAACTGACTGTGAGGAGAGAGATGAGCAGGACCTCGAG 480
  
```

QY 1123 ttccagtgcc--tgagagaagacagccagtgctgtgaaaggggcctgtgttcacgt 1180
 |||||||
 DB 481 TTCCAGTGGCTCTGCAGAGAGAGACAGGCGAGTCTGGAAGGGGCGCTGTGCTTACGT 540
 |||||||
 QY 1181 tgcatacctgaa 1193
 |||||||
 DB 541 TCCATGACCTGAA 553

RESULT 8
 BE813148/C
 LOCUS
 DEFINITION RC1-BN0005-190700-017-h02 BN0005 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE813148
 VERSION BE813148.1 GI:10245382
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 594)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,R.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20020263
 CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=612-RC1-BN0005-190>)
 700-017-h02et3-2000-07-19et4-1)
 Seq primer: puc 18 forward
 High quality sequence stop: 527.

FEATURES
 source
 1..594
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BN0005"
 /dev_stage="Adult"
 /note="Organ: breast_normal; Vector: puc18; Site: 1: SmaI;
 Site: 2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. letters patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 123 a 161 c 177 g 133 t
 ORIGIN

Query Match 26.3%; Score 513.8; DB 136; Length 594;
 Best Local Similarity 95.6%; Pred. No. 1.7e-118;
 Matches 561; Conservative 0; Mismatches 22; Indels 4; Gaps 3;

```

QY 1284 tggccccccttgatgagcattcaagagagaggggtggtgaagaagaatagtgctt 1343
    |||||||
DB 594 TGGCCCCCTGGAGTGCATTCAGAGAGAGATAGCTCTGGGTGAAGAAATATTTGT 535
    |||||||
QY 1344 gaatcttcttgaaagcgtcaaggacccccgcccacatctcctggaacgtcaacgg 1403
    |||||||
  
```

[illegible]

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3877338"
/clone_id="NTH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pCMV-Sport6; Site:1: NotI; Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT

```

Average insert size 1.8 Kb.		Library constructed by Life Technologies.	
BASE COUNT	158 a 218 c 196 g 143 t		
ORIGIN			
Query Match	23.9%	Score 466.4	DB 135; Length 715;
Best Local Similarity	99.8%	Pred. No. 1.5e-166	
Matches 467; Conservative	0; Mismatches	1; Indels	0; Gaps

QY	1448	ccctgaagcttcctcgtgacccccgagcgtgtctggagaacaggtgttgaatgcacgacctca	1507
Db	1	ccctgaatgtctctctgtgacccccgagcgtgttggagacaggtttgaatgcacagggcttcca	60
QY	1508	acgacctgggcaaaaacaccacgacalcctctcttcctgtagcgtgtcaatttaaccacctca	1567
Db	61	acgacctgggcaaaaacaccacgacalcctctcttcctgtagcgtgtcaatttaaccacctca	120
QY	1568	caccagatctccacacacacccactggtccctcaagcactccacgtccagctctctataccag	1627
Db	121	caccagactccacacacacccactggtccctcaagcactccacgtccagctctctataccag	180
QY	1628	ccaacagcaccctccacacacagagaagaagcgtccgagccgagagacccgaggtgtgcatcg	1687
Db	181	ccaacagcaccctccacacacagagaagaagcgtccgagacccgagagacccgaggtgtgcatcg	240
QY	1688	tggctgtgattgtgtgcatcctcgtgtccgtgagtgctgtgagcgtgtcctctatctctct	1747
Db	241	tggctgtgattgtgtgcatcctcgtgtccgtgagtgctgtgagcgtgtcctctatctctct	300
QY	1748	ataagaagggtcaagctgtccgtgtcagaagcgtccagaggaacagagagatacagctgtcccg	1807
Db	301	ataagaagggtcaagctgtccgtgtcagaagcgtccagaggaacagagagatacagctgtcccg	360
QY	1808	cttcgcaagagcgaactctgtagtgaagttaagctcagataaagctccacagaagatgtggcc	1867
Db	361	cttcgcaagagcgaactctgtagtgaagttaagctcagataaagctccacagaagatgtggcc	420
QY	1868	tcttcgcaaggtcagcagcgtgtgacaagaagggtccggtgagagaccagagag	1915
Db	421	tcttcgcaaggtcagcagcgtgtgacaagaagggtccggtgagagaccagagag	468
RESULT 10			
AA983120			
LOCUS	AA983120	729 bp	mRNA
DEFINITION	ua35d07.r1 Soares_mammary_gland_NbMNG Mus musculus cDNA clone		EST
	IMAGE:1348717 5' similar to SW:0018_HUMAN P43121 CELL SURFACE		27-MAY-1998
	GLYCOPROTEIN MUC18 PRECURSOR ; mRNA sequence.		
ACCESSION	AA983120		
VERSION	AA983120.1	GI:3160398	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 729)		
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubugue,T.,		
	Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,		
	Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,		
	Waters,J.B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and		
	Waterston,R.		
TITLE	The Mashu-HMM Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	Contact: Marra M/Mouse EST Project		

|||||
Db 200 GACTGTTTCTTGTCCACAGAGAACCAACATCTTCCGGGTGGGAGGCGCAG 259
|||
Qy 241 ggcagagcaaacctgaggagtaagacagcagcctcagcagcagagagctact 300
|||||
Db 260 ggcagagagagcctgaggagtaagacagcagcctcagcagcagagagctact 319
|||
Qy 301 ctcgagcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360
|||||
Db 320 ctcgagcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 379
|||
Qy 361 ctcgagcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420
|||||
Db 380 ctcgagcctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 439
|||
Qy 421 atccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 480
|||||
Db 440 atccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 499
|||
Qy 481 tctgtgag 540
|||||
Db 500 tctgtgag 559
|||
Qy 541 ctgaa 545
|||
Db 560 CTAA 564

RESULT 12

BE226518 602 bp mRNA EST 06-JUL-2000
LOCUS BE226518
DEFINITION la21g07.y1 Mouse E10 5 12 5 Pancreas cDNA Library Mus musculus cDNA
5' similar to TR-095812 095812 CELL SURFACE GLYCOPROTEIN P1H12
PRECURSOR ;, mRNA sequence.

ACCESSION BE226518
VERSION BE226518
KEYWORDS GI:8931754
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 602)
AUTHORS Melton, D., Meadows, A., Clifton, S., Hillier, L., Marra, M., Pape, D.,
Wyllie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter,
E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole,
R., Tsagaris, H., Williams, T., Jackson, Y., and Bowers, Y.
WashU-Harvard Pancreas EST Project
Unpublished (2000)
Other ESTs: la21g07.x1

TITLE JOURNAL
COMMENT Contact: Douglas Melton
WashU-Harvard Pancreas EST Project
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu, meadows@fas.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Adimika Meadows
(meadows@fas.harvard.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 454.
Location/Qualifiers

FEATURES

SOURCE

1. 602
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
/clone_lib="Mouse E10 5 12 5 Pancreas cDNA library"
/sex="Both"
/tissue_type="Pancreatic Bud"
/dev_stage="Embryonic day 10.5 and 12.5, mixed"
/lab_host="DH10B"

/note="Organ: Pancreas; Vector: pSPOR1; Site_1: Not 1;
Site_2: Sal 1; Library constructed using Superscript
Plasmid Library kit (Life Technologies). cDNA made by
oligo-dT priming. Size-selected by column fractionation;
average insert size 1.47 kb. Primary library, unamplified.
cDNA Library Preparation: Goulin Chen."

BASE COUNT 157 a 166 c 147 g 132 t
ORIGIN

Query Match 22.3%; Score 435.2; DB 105; Length 602;
Best Local Similarity 82.8%; Pred. No. 9,7e-99;
Matches 497; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
Qy 1339 gctgtgaatcgtctctgtgtgaagcgtcagagaccccgccacccatctctgtgaagc 1398
|||
Db 3 GTGCTGANTCTGTTTGTGAGGCTTGAGACATCTCAGCCACATCTCTGTGAAATGTC 62
|||
Qy 1399 aacgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1458
|||
Db 63 AATGTTGCGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 122
|||
Qy 1459 ctctgtgaccccgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1518
|||
Db 123 CTGTGAGCGGAGAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 182
|||
Qy 1519 aaaaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1578
|||
Db 183 TCMAACACACACACACACATTTGTTGATGATGATGATGATGATGATGATGATGAT 242
|||
Qy 1579 aacacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1638
|||
Db 243 AGCCAAAC 302
|||
Qy 1639 tccacag 1698
|||
Db 303 TCCACAG 362
|||
Qy 1699 gctgtcaccctgt 1758
|||
Db 363 GTGTGTACCTTGT 422
|||
Qy 1739 aagctgt 1818
|||
Db 423 AAGCTGCGATGT 482
|||
Qy 1819 gaactgt 1878
|||
Db 483 GAATGT 542
|||
Qy 1879 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1938
|||
Db 543 ATCAACTGT 602
|||

RESULT 13

AW611108 642 bp mRNA EST 23-MAR-2000
LOCUS AW611108
DEFINITION un77a02.y1 Sugano mouse kidney m1a Mus musculus cDNA clone
IMAGE:2536874 5' similar to SW:W018.HUMAN P43121 CELL SURFACE
GLYCOPROTEIN MUC18 PRECURSOR ;, mRNA sequence.

ACCESSION AW611108
VERSION AW611108.1
KEYWORDS GI:7315849
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 642)
AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wyllie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person,
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter,
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,

Query Match	21.7%:	Score 422.6:	DB 93:	Length 642:
Best Local Similarity	79.0%:	Pred. No. 1.4e-95:		
Matches 503:	Conservative 0:	Mismatches 134:	Indels 0:	Gaps 0:
QY	294	ggctactctgagccctgagactcaagtcacaccccaagaagagcgatctctctgtgccaggg	353	
DB	6	GGCTACTCTGGCCCTGAGTCACTGTCACATCCCCAGATGAGACGAATGTTCTGTGTAAGG	65	
QY	354	caagagccctcggtgcccaaggaatccgcatccagctccgcttacaagaagctccggagga	413	
DB	66	CAGAAGACACACGGCTCCAGATCACTACGTTGAGCCYTCAGGTTCCTAAAGCCCCAGAGGA	125	
QY	414	ggcaaacatccaggtcaaccccccgaggatccctctgttaaacagtaagaagccctgaagagt	473	
DB	126	ACCACACTATTCAAGCAATGTCTGTGGGCATTCATGTGTGACAGGCAAGACTTAGGGAGGT	185	
QY	474	cgctccctcggttagggaggaagcgggttaeccatctcctcaagtcattctgttacaagaatgg	533	
DB	186	TGCTACCTGTGTGGGAAATAACGGCTACCCCATTCCTCAAGTCTCTATGTGTAACAAAGACG	245	
QY	534	ccgagctctgaaggaggaagaacccgggtccacatccatgaagtcgtcccaagctgttgatc	593	
DB	246	TCTGCGCTTGGAAGAGGAGAGAAACCGAGTTCAATTCACGTATCATCAGCATTTGTGGAAGTC	305	
QY	594	gagtggtttgtaaccctctgcagagatattctgaagcgacacagctggtttaaagaagacaaaga	653	
DB	306	CAGTGGCTTGACACCTTGGAAGAGTGTCTGCACTGCACCCCTAGTTAAAGAGACAAAGA	365	
QY	654	tggccagtttactgttagctcaactcaacccggctccggtccagttgggaacaaatagaagatc	713	
DB	366	TGCCAGTTTACTGTGACATCAGCTACGCTACCGCTACCCCATGTGGGAACCATATAGAGGAATTC	425	

Query Match	Similarity	21.0%	Score 410	DB 142	Length 414
Best Local Similarity	99.0%	Pred. No. 1	9c-92		
Matches 410	Conservative	0	Mismatches 4	Indels 0	Gaps 0
Qy	744	gacagaaagaagtgctgctgaagtgagcccgctgggaatcgtcaagaaaggagccgct	803		
Db	1	GACAGAAAAGTGTGCTGCTGAATGGAGCCCGCTGGGAATGCTCAAGGAAGGGGACCCGCT	60		

QY 804 ggaatcaggtgttctgctgctgcaacccctccacacacttcagcatcagcaagcagaa 863
 |||||||
 DB 61 GGAATTCAGGTGTGGCTGTATGCAACCCCTCCACACTTGAGCTCAGCAAGCAAA 120
 |||||||
 QY 864 ccccgagcagggagagagagagagagacacacacacagaggggtccctggtgca 923
 |||||||
 DB 121 CCNCGACACCGAGGAGAGAGAGAGAGACACACAGCAGCGGGTCTGGTGGTGA 180
 |||||||
 QY 924 gcttgcgcgaaggaacacacagtggcgtatgaatgtcaggcctggaacttgacacat 983
 |||||||
 DB 181 GCTGCTCCGGAAGGAAACACTGGCGCTATGATGTNAGGGCTTGACTTGGACACCAT 240
 |||||||
 QY 984 gatacagctgtcgtgagtgagcaccaggaactactgtgtaactatgtgtcagctccagat 1043
 |||||||
 DB 241 GATATCGCTGCTGAGTGAACCAAGAGACTACTGTGTAATGTGTCTACGTCCGANT 300
 |||||||
 QY 1044 ggttcgcgaagcagccctcgtgagagacaggaagcagacagctccacccctgagcagaga 1103
 |||||||
 DB 301 GAGTCCCGCAGCCCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
 |||||||
 QY 1104 ggttagcagagacagctcagctcagctgagctgagagagagagagagagagagagagag 1157
 |||||||
 DB 361 GAGTAGCCAGGACCTCGAGTTCAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 414
 |||||||

RESULT 15

AA191148 423 bp mRNA EST 10-MAR-1998
 LOCUS zpb6e01.r1 Stratagene HeLa cell s3 937216 Homo sapiens cDNA clone
 DEFINITION IMAGE:627096 5' similar to SW:MU18_HUMAN P43121 CELL SURFACE
 GLYCOPROTEIN MUC18 PRECURSOR ; mRNA sequence.
 AA191148
 VERSION AA191148.1 GI:1779842
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 423)
 HILLIER, L., ALLEN, M., Bowles, L., Dubouche, T., Gelsel, G., Jost, S.,
 Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Mottin,
 J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theisling, B.,
 White, T., Wyllie, T., Waterston, R. and Wilson, R.
 WashU-NCI human EST Project
 Unpublished (1997)
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 2045 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 355.
 Location/Qualifiers
 1..423
 /organism="Homo sapiens"
 /db_xref="GDB:5045794"
 /db_xref="taxon:9606"
 /clone="IMAGE:627096"
 /clone_1ib="Stratagene HeLa cell s3 937216"
 /sex="female"
 /dev_stage="HeLa S3 cell line"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; Cloned unidirectionally. Primer: Oligo dT. HeLa S3
 epitheloid carcinoma cells grown to semi-confluency
 without induction. Average insert size: 1.5 kb; Uni-ZAP XR
 vector. ~5' adaptor sequence: 5' GAATTCGGCAGAG 3' ~3'
 adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

FEATURES

source

BASE COUNT

98 a 137 c 108 g 80 t

ORIGIN

Query Match 20.5%; Score 399; DB 3; Length 423;
 Best Local Similarity 99.3%; Fred. No. 1.1e-89;
 Matches 421; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
 QY 1342 ttgaatctgtctgttgaagcgtcagagcaccgcccacacatctccttgaagcgtcaac 1401
 |||||||
 DB 1 TTGAATCTGCTGTGTGAAGCGTCAAGGCGACCCCGGCCACATCTCTGGAACGTCAAC 60
 |||||||
 QY 1402 ggcacgcaagtgacaacaagaacacacacacagcagcagcttgagca -ccctgaatgtcct 1460
 |||||||
 DB 61 GGCACGGCAGCAAGTGAACAAGAACCAAGATCCACAGGAGTCTTGAGCACCCCTGATGTCT 120
 |||||||
 QY 1461 cgtgaccccgagagcgtgttgaagacaggtgttgaatgaacagcctccaaagcagcagaa 1520
 |||||||
 DB 121 GGTGACCCCGAGAGCTGTTGGAGACAGAGTGTGAATGACAGGCTCCACAGCACTGGGCAA 180
 |||||||
 QY 1521 aaacacagcatcctctcctgtgagctggtcaatttaacacaccccaacacagactccaa 1580
 |||||||
 DB 181 AATACACGATCTCTCTCTGAGAGTGTCAATTTAACCAACCTCACACAGACTCCAA 240
 |||||||
 QY 1581 cacaacacatggtcctcagcaacttcacagcagcagcagcagcagcagcagcagcagcagc 1640
 |||||||
 DB 241 CACAAACCACTGCTCAGCACTTCACACTGCACTGCTCATACAGAGCCACAGACCTC 300
 |||||||
 QY 1641 cagagagagaaagctgtccgagcagagcagagcagagcag -tgtcaatcgtgtgtatgt 1699
 |||||||
 DB 301 CACAGAGAGAAAGCTGCGGAGCGGAGAGAGCGGGGCTGTGATGTGTGATTG 360
 |||||||
 QY 1700 tgtgaatccctgtcctcgtgagcgtgtgagcgtgtcctctatcttccctataaagaagagca 1759
 |||||||
 DB 361 TGTGATCTCTGCTGCTGCGGCTGTGCTGTGCTGTCTATTTCTCTATTAAGAAGGGCA 420
 |||||||
 QY 1760 agc 1762
 ||||
 DB 421 AGC 423

Search completed: March 26, 2001, 06:56:08
 Job time: 4484 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: March 23, 2001, 12:30:37 ; Search time 19.45 seconds
(without alignments)
1135.691 Million cell updates/sec

Title: US-09-653-961-2
Perfect score: 3363
Sequence: 1 MGLPVLCAFLLAACCCPR.....SSGDKRAPDQGEKIDLRH 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_36:*
1: /SIDSL/gcgcdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDSL/gcgcdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDSL/gcgcdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDSL/gcgcdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDSL/gcgcdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDSL/gcgcdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDSL/gcgcdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDSL/gcgcdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDSL/gcgcdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDSL/gcgcdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDSL/gcgcdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDSL/gcgcdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDSL/gcgcdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDSL/gcgcdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDSL/gcgcdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDSL/gcgcdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDSL/gcgcdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDSL/gcgcdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDSL/gcgcdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDSL/gcgcdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDSL/gcgcdata/geneseq/geneseqp/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3341	99.3	646	20	Y42404
2	3327	98.9	646	20	Y42405
3	494.5	14.7	583	17	R97230
4	492.5	14.6	583	17	W06891
5	492.5	14.6	583	19	W47088
6	466	13.9	570	17	R97231
7	266	7.9	3117	21	Y53667
8	252	7.5	1380	20	Y08402
9	244.5	7.3	1381	20	Y13564
10	244.5	7.3	647	14	R39686
11	237.5	7.1	736	14	R39682
12	236	7.0	4412	21	Y53666

13	234	7.0	848	21	Y88565	Human NCAM 140KD 1
14	233	6.9	647	11	R08117	Vascular cell adhe
15	233	6.9	647	14	R38549	VCAM-6D. Syntheti
16	233	6.9	647	21	Y59498	Human VCAM-1 prote
17	233	6.9	1496	20	W81030	Melanoma associate
18	233	6.9	1496	21	W70469	Human p53 target m
19	231.5	6.9	753	20	W83927	Human t85 protein.
20	231.5	6.9	828	12	R13905	Rabbit ATHERO-ELAM
21	230	6.8	735	14	R38550	VCAM/ICAM-1. Synt
22	230	6.8	739	12	R08118	Vascular cell adhe
23	230	6.8	739	13	R10316	1E7/267 staloglyco
24	230	6.8	739	14	R21081	VCAM variant with
25	230	6.8	739	15	R38548	Sequence of the 1E
26	230	6.8	739	15	R49733	Amino acid sequenc
27	230	6.8	739	19	W46734	Human VCAM-1. Hom
28	228	6.8	702	10	P94014	Human VCAM-1. Hom
29	227	6.7	739	12	R13906	Human VCAM-1. Hom
30	226	6.7	738	12	R13251	Human VCAM-1. Hom
31	226	6.7	1257	20	W74152	PECAM-1. Homo sap
32	225.5	6.7	698	16	R65168	PECAM-1. Homo sap
33	225	6.7	698	16	W22844	Human L1 cell adhe
34	225	6.7	734	17	W00182	Human carcinoembry
35	225	6.7	1070	18	W08747	Human carcinoembry
36	224.5	6.7	1649	20	Y08404	Human colon carcino
37	224.5	6.7	702	10	P93999	Human Robo1 protei
38	224.5	6.7	702	15	P54713	Human Robo1 polyp
39	224	6.7	702	17	W06872	Amino acid sequenc
40	224	6.7	702	20	W83137	Carcinoembryonic a
41	224	6.7	702	20	W83137	Carcinoembryonic a
42	224	6.7	644	14	R39685	CEA protein. Homo
43	222.5	6.6	698	9	R81229	VCAM-6D/ICAM-2. S
44	222	6.6	702	9	P81222	Carcinoembryonic a
45	222	6.6	702	9	P81222	Carcinoembryonic a

ALIGNMENTS

RESULT 1	Y42404	standard; Protein; 646 AA.
ID	Y42404	
AC	Y42404	
DT	01-DEC-1999	(first entry)
XX		
DE	Human MUC18 amino acid sequence.	
XX		
KW	prostate cancer; melanoma; cell adhesion; glycoprotein;	
KM	metastasis; treatment; detection; diagnostic test.	
XX		
OS	Homo sapiens.	
XX		
PN	W09945392-A1.	
XX		
PD	10-SEP-1999.	
XX		
PE	02-MAR-1999; 99WC-US04850.	
XX		
PR	03-MAR-1998; 98US-0076664.	
XX		
PA	(UYEM-) UNIV EMORY.	
XX		
PI	WU G.	
XX		
DR	NPI; 1999-540899/45.	
XX		
DR	N-PSDB; Z20930.	
XX		
PT	Detection of metastatic prostate cancer, by detection of MUC18	
XX		
PT	expression in prostate cancer cells	
XX		
PS	VCAM-6D/ICAM-2. Synt	
XX		
PS	Claim 17; Page 7; 80pp; English.	
XX		

CC This is the amino acid sequence of the Human MUC18 cell adhesion
CC glycoprotein, which is expressed on the surface of melanoma cells.
CC and can be used as a marker for prostate cancer.
CC This MUC18 (hMUC18) cDNA sequence is different from the hMUC18 cDNA
CC given in Genebank accession number N28882, 220931. The deduced amino acid
CC residues from this cDNA are very different from that published by
CC Johnson's group in 1989.
CC The presence of this glycoprotein has been correlated with the ability of
CC melanomas to metastasize. MUC18 is also associated with normal vascular
CC tissue, and on the smooth muscle of venules, and it expresses
CC sporadically on capillary epithelium.
CC The method can be used as a diagnostic test for prostate cancer which
CC has a relatively high potential for metastasis or which has metastasized.
CC The physician can then choose the appropriate surgical, chemotherapeutic
CC or radiation treatment regimens. In addition, antibody specific to MUC18
CC can be used to prevent metastasis of PCCs.

XX
XX
Sequence 646 AA:

Query Match 99.3%; Score 3341; DB 20; Length 646;
Best Local Similarity 99.5%; Pred. No. 2,7e-227;
Matches 643; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGLPLVCAFLIAACCCPRVAGVGEADQAPALVEVEVSTALLKGLSOSGNSHV 60
DB 1 mglplvcafliaacccprvagvgpapelvevestallkcglsqsgnshv 60
QY 61 DMFSVHKERKTLIFRVROGOGSEGEYEQRLSLDRGATLALNQVPODERIFLCOGKR 120
DB 61 dmfsvhkerkrlifrvrgogogsegeyegrlslqdrqataltqvpqderiflcgkr 120
QY 121 PRSQEYRIQLRYKKAPEEPNIQVNLGIPIVNSKEPEEVAATCGRNGYPIQVYWKNGRP 180
DB 121 prsqeyriqlrykkapeepniqvnlgipivnskepeevaatcvgngypipqviwkngrp 180
QY 181 LKEEKNRVHIQSSQTVESGGLYTLQSLIKAOVKEDKDAOFYCELNRYLPSGNHMKESRE 240
DB 181 lkeeknrvihsqtvessgyltqlslkqavlkedkdaofyccelnrylpsgnhmkesre 240
QY 241 VIVPVFYPEKRWLKEVEPVGMLEKGDRIEIRCLADGNPPIHFSISKONPSTREAEEETT 300
DB 241 vivpvfypekrlwleevpvgmlkegdrieircladgnppihfsiskpnstreaeeett 300
QY 301 DNGVNLVLEPAKREHSGRYCOGLDIDTMTSLISEQELLVNVSDVRYSPAPEROEGSS 360
DB 301 dngvnlvleparkehsgryecogldidmtslisepellvnyvsdvryspaeperegss 360
QY 361 LFTLTCAESSQDLFEQMLEEETGQVLEKRGVLIQLMDLKEAGGVCYASVPSIDGLNRT 420
DB 361 lftltcaessqdlfeqmwleeeegqvleirgvlqhlkreaaggyrcvasvpsidglnrt 420
QY 421 QLVNVAIFGPPMAFKERKRWAKENNVLMSCFASGHPPTTSMWNGTASQDDPOKV 480
DB 421 qlvnvaifgppmafkerywvkenmvllmscfasghprrtllsmwngtaseqddpqkv 480
QY 481 LSTLVNLPVPELLETGVECTASNDLGKNTSILELVNLTTLTTPDSNTTGLSTASPH 540
DB 481 lsltnlvlpelletgvectasnldgkntsilflvnltnltltpdsnttglstasph 540
QY 541 TRANSTTERKRLPEPSRCGVIVAVTCILVAVLIGAVLYFKKGLPCRRSGKOETL 600
DB 541 transtterkrlpepsrgvivaavtcilvavligavilykkglpcrrsgkoetl 600
QY 601 PPSRKSELVVEVSKDLPEMGLLOSGGKRAKPGOGKEVYIDL 646
DB 601 ppsrkselvvvevskdklpeemgllgssgdkrapdgqekyidl 646

RESULT 2
Y42405
ID Y42405 standard; Protein: 646 AA.
XX

AC Y42405;
XX
DT 01-DEC-1999 (first entry)
XX
XX Human MUC18 amino acid sequence. (Johnson 1989).
DE
XX
XX prostate cancer; melanoma; cell adhesion; glycoprotein;
KW metastasis; treatment; detection; diagnostic test.
XX
XX Homo sapiens.
OS
XX
XX W0945392-A1.
XX
XX 10-SEP-1999.
PD
XX
XX 02-MAR-1999; 99WO-US04850.
PF
XX
XX 03-MAR-1998; 98US-0076664.
PR
XX
XX (UYEM-) UNIV EMORY.
PA
XX
XX Wu G;
PI
XX
XX WPI; 1999-540899/45.
DR
XX
XX N-PSDB; 220931.
PT
XX
XX Detection of metastatic prostate cancer, by detection of MUC18
PT expression in prostate cancer cells -
PS
XX
XX disclosure; Page 8; 80pp; English.
XX
XX This is the amino acid sequence of the Human MUC18 cell adhesion
CC glycoprotein, which is expressed on the surface of melanoma cells,
CC and can be used as a marker for prostate cancer.
CC This sequence was originally published in 1989 by Johnson et al
CC (Genebank N28882).
CC This sequence differs from the one proposed by the inventors, this amino
CC acid sequence is 603 amino acid residues while the present sequence
CC 220930 has 646 amino acid residues.
CC The presence of this glycoprotein has been correlated with the ability of
CC melanomas to metastasize. MUC18 is also associated with normal vascular
CC tissue, and on the smooth muscle of venules, and it expresses
CC sporadically on capillary epithelium.
CC The method can be used as a diagnostic test for prostate cancer which
CC has a relatively high potential for metastasis or which has metastasized.
CC The physician can then choose the appropriate surgical, chemotherapeutic
CC or radiation treatment regimens. In addition, antibody specific to MUC18
CC can be used to prevent metastasis of PCCs.

XX
XX
Sequence 646 AA:

Query Match 98.9%; Score 3327; DB 20; Length 646;
Best Local Similarity 98.9%; Pred. No. 2,6e-226;
Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MGLPLVCAFLIAACCCPRVAGVGEADQAPALVEVEVSTALLKGLSOSGNSHV 60
DB 1 mglplvcafliaacccprvagvgpapelvevestallkcglsqsgnshv 60
QY 61 DMFSVHKERKTLIFRVROGOGSEGEYEQRLSLDRGATLALNQVPODERIFLCOGKR 120
DB 61 dmfsvhkerkrlifrvrgogogsegeyegrlslqdrqataltqvpqderiflcgkr 120
QY 121 PRSQEYRIQLRYKKAPEEPNIQVNLGIPIVNSKEPEEVAATCGRNGYPIQVYWKNGRP 180
DB 121 prsqeyriqlrykkapeepniqvnlgipivnskepeevaatcvgngypipqviwkngrp 180
QY 181 LKEEKNRVHIQSSQTVESGGLYTLQSLIKAOVKEDKDAOFYCELNRYLPSGNHMKESRE 240
DB 181 lkeeknrvihsqtvessgyltqlslkqavlkedkdaofyccelnrylpsgnhmkesre 240
QY 241 VIVPVFYPEKRWLKEVEPVGMLEKGDRIEIRCLADGNPPIHFSISKONPSTREAEEETT 300
DB 241 vivpvfypekrlwleevpvgmlkegdrieircladgnppihfsiskpnstreaeeett 300

```

Db      241 vtyvfyfptekvwleevygmikegdvrelrcladgnpphfsisknpstreaeetn 300
QY      301 DNGVLVEPARKHSGRYECQGLDPTMTISLSEPOELVNVYSDVRSAPAREOGSS 360
Db      301 dmyvlvleparkehsgrycqawltmtisllsepeellvnyvsvdrvsapaerqegss 360
QY      361 LTLTCEAESSODLEFQMLREETQVLERGPVLOLHDLKREAGGVCVAVSPISPELNT 420
Db      361 lltlceaessdlefqmwleedqvlergvpvlqhlkreaaggvrcvaspslpglnt 420
QY      421 QLVNVALFGPPMAFKERKWKVAKENVNLSCESAGHPPTISWNVNCTASBDQDPQRY 480
Db      421 qlyklsalfgppmafkerkwvkenmvnlscsaaghpptlismvngfasedqdpqrv 480
QY      481 LSTNLVNLVPELLETCGVECTASNDGKNTSILELVNLTITLTPDSNTTGTSTASPH 540
Db      481 lstnlvnlvpeellectgvectasndlgkntslllelvnltiltpdsntltgtstasph 540
QY      541 TRANSTSTERKLPPEPSRGVIVAVICILVAVLGAVLFLYKKGKLPGRSGKOETL 600
Db      541 tranststerklpepsrgvivaivcillvavlgavlyflykkgklpgrtsqkgetl 600
QY      601 PPSRKELVVEKSDKLPPEMGLGSSGDKRAPGDGGERKYLDRH 646
Db      601 psrkelyvveksdklppeemglgssgdkrapgdggekylldrh 646

RESULT 3
R97230
ID      R97230 standard; Protein: 583 AA.
XX
AC      R97230;
XX
DT      08-OCT-1996 (first entry)
XX
DE      Stem cell marker HCAPro.1.
XX
KW      Stem cell marker; haematopoietic cell antigen; HCA; HCASeq.1;
KM      HCAPro.1.
XX
OS      Homo sapiens.
XX
PN      EP16146-A2.
XX
PD      12-JUN-1996.
XX
PF      02-FEB-1995; 95EP-0300661.
XX
PR      06-DEC-1994; 94US-0352323.
XX
PA      (SYST-) SYSTEMIX INC.
XX
PI      Gearing D, Uchida N, Yang Z;
XX
DR      WPI: 1996-269990/28.
XX
DR      N-PSDB: T28819.
XX
DE      DNA encoding human haematopoietic cell antigen proteins - useful as
PT      stem cell marker proteins in functional studies and for antibody
PR      prodn.
XX
PS      Claim 8; Fig 4A; 17pp; English.
XX
CC      HCAPro.1 (R97230) and HCAPro.2 (R97231) are novel human
CC      haematopoietic cell antigens associated with stem cells. They
CC      are the respective products of nucleic acid sequences HCASeq.1
CC      (T28819) and HCASeq.2 (T28820) obtd. from human CD34+ bone marrow.
CC      Expression systems are provided for prodn. of recombinant HCA
CC      proteins. The proteins are useful as stem cell marker proteins in
CC      functional studies, and can also be used to produce antibodies in
CC      that allow the purification of stem cells from haematopoietic and
CC      other sources.

```

```

XX      XX      Sequence      583 AA:
SQ      SQ
Query Match      14.7%; Score 494.5; DB: 17; Length 583;
Best Local Similarity 24.8%; Pred. No. 6.5e-27;
Matches 145; Conservative 121; Mismatches 256; Indels 63; Gaps 23;

QY      41 GSTALLKCGLSOGQNLSHVDFSVYKKEKRTLIRVRGOGQS----EPGEYEQRLSD 96
Db      36 gdtllipcrldvpg-nlmfgkwnyekpdpqspvllaifstksqyddvpeyadrlhse 94
QY      97 RGATLALQVTPQDERIRFLCO-GKPKSQEYRIOLRYKAPDEPNIOVNLGIPVNSKEP 155
Db      95 -nyltsisnarnsdckrtvcmlvtedhvfearlvkfkqpskpelvskalfi--eteql 151
QY      156 EEWATCVGRNGVPIPOVIVYKNGRPLKEEKRRVHIOSSQYVES-SGLYTLQSTLKAQLV 214
Db      152 kklgdci sedsyppdgnltwryngkvlnpliegavvllfkemdpvtqilymtstleytk 211
QY      215 EDKDAQFYCELNRYLPSGNHMKESREYTVPEYFTEKRWLEV-EPVGMLEKGRVEIRCL 273
Db      212 adlqmpficsvtylygpsqgkclihsegavfdiylpyleqvltqvlppknaikegdnitlcl 271
QY      274 ADGNPPRH--FSISKQNPSTRAEETNDNGVLYLEPARKHSGRYECQGLDPTMTIS 330
Db      272 gngppppeeelflylpgperrlrsntytltd-----vyrnatgdykcslldkksmia 323
QY      331 LLSPEQELVNVYSDVRSAPAP-EROEGSSLTITLCEAESSODLEFQMLREETQVLERG 389
Db      324 sta-----ltvhyt-dlslpsgevtrtqgdalvpsclsisrnatlvvmkdnlr--lrss 376
QY      390 PVLOLHDLKREAGGGRVCAVSPSIPGILNRTOLVNVAFPGPMMAFKERKWKVAKENVLN 449
Db      377 p--sfsslhqdagnyucetalgveglkkresitllivegkp--qlkmcktktdpslstk 432
QY      450 LSCESAGHPPTISWNVNCTAS---EODDQPO---RYLSTNLVNLVPELLETCGVECTAS 503
Db      433 lchvegfpkpaqlwtlqgsavlnqleespylngrtys--kllispeenvt-ltctaen 489
QY      504 DUGKNTSILELVNLTTLTPDSNTTGTSTASPHSTRANSTSTERKLPPEPSRGVIV 563
Db      490 qlertvnslnvsalsi-----pe-----hdeadeidsentr-ekvndgklliv 530
QY      564 AVIYCIILVAVLAVLYFLY-KKGKLPGRSGKOETLTPPSRSE 607
Db      531 givvglllaalvagvywlymkkstaskhvnkdlgmneenkkle 575

RESULT 4
W06891
ID      W06891 standard; Protein: 583 AA.
XX
AC      W06891;
XX
DT      18-MAR-1997 (first entry)
XX
DE      Human activated leukocyte-cell adhesion molecule ALCAM.
XX
KW      Activated leukocyte-cell adhesion molecule; ALCAM; CD6 ligand;
KM      thymic epithelial cell; inflammation; allograft rejection;
KM      neurodegenerative disease.
XX
OS      Homo sapiens.
XX
XX
FH      Key      Location/Qualifiers
FT      Peptide      1..27
FT      Protein      28..583
FT      Domain      28..527
FT      /label= "Extracellular domain
FT      /note= "the extracellular domain can be subdivid"

```


FT	Domain	538..551	/label= "Transmembrane_domain"
FT	Domain	552..583	/label= "Extracellular_domain"
FT	Modified-site	91	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Modified-site	95	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Modified-site	167	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Modified-site	167	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Misc-difference	258	/label= Asn, Ser /note= "variation owing to polymorphism in HL60 and T-cell derived clones"
FT	Modified-site	265	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Misc-difference	301	/label= Met, Thr /note= "variation owing to polymorphism in HL60 and T-cell derived clones"
FT	Modified-site	306	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Modified-site	337	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Modified-site	361	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Modified-site	457	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Modified-site	480	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Modified-site	499	/label= "Glycosylation /note= "putative N-glycosylation site"
FT	Modified-site	499	/label= "Glycosylation /note= "putative N-glycosylation site"
PN	W09634880-A1.		
XX	07-NOV-1996.		
XX	29-APR-1996;	96WO-US06010.	
XX	01-MAY-1995;	95US-0432016.	
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.		
PA	(UTDU-) UNIV DUKE.		
PI	Aruffo A, Bowen MA, Haynes BF, Marguardt H, Patel D;		
DR	WPI: 1996-506097/50.		
XX	N-PSDB: T46075.		
XX	CD6 ligands present on surface of thymic epithelial cells - used to develop prods. for treating e.g. inflammation, organ allograft rejection or neurodegenerative diseases		
PS	Example 4: Fig 29; 112pp; English.		
CC	Human activated leukocyte cell adhesion molecule (W06891) or ALCAM is a CD6 ligand present on the surface of thymic epithelial cells, monocytes, activated T-cells, and other cells. Its amino acid sequence was deduced from cDNA clones (T46075) obtd. from HL60 and T-cell lines. ALCAM polypeptides can be produced, e.g. as a fusion proteins, in transformed host cells. They are useful for inhibiting CD6/CD6 interactions between cells and to screen for agents that modulate this interaction. Inhibitors can be used to		

Query Match	14.6%;	Score 492.5;	DB: 17;	Length 583;
Best Local Similarity	24.8%;	Pred. No. 9e-77;		
Matches 145;	Conservative 121;	Mismatches 256;	Indels 63;	Gaps 23;
CC treat e.o. inflammation, multiple sclerosis, inflammatory uveitis, rheumatoid arthritis, T-cell mediated vasculitis syndromes, organ allograft rejection and neurodegenerative diseases.				
XX	Sequence	583 AA;		
SO				
QY	41	GS7LLKCGISQSGNLSHVDWFSVVKERTLIFRVROGQGS----	EPGEYEQRLSIOD	96
DB	36	gdtlllporldvbp-nlmfgkwyekpdpypfiafrssktsvgyddvpeykdrlllse		94
QY	97	RGATATLQVTPQPERFLFCQ--EKRRSEGYEIOLRVKAKEEPIQNPGLIPVNSKEP	155	
DB	95	-nytlslsnarisdskrfvcmlvtelnvteapllkvkfkqskpeylskafll--eteqj	151	
QY	156	EEVATCYGRNGYPIPOVIWTKNGRPLKEKNRNVHIQSSQYES--SGLTLLDSILKADLV	214	
DB	152	kllgdcsedpsypgnaltwyrngkvlhpldegavvllfkemdpvqlytmsltleytkc	211	
QY	215	EDKDAQFYCELNVRLPSGNHMKESREVTVPVPEYKEWLEV--EPVGMLKEGDEVEIRCL	273	
DB	212	adiqmflfcsvctlygpgsgkqlhseqavfdiyyrceqvrlqylppkxalkgegnhltkl	271	
QY	274	ADGNPPPH--FSISKONPSTREAEEETTNDNGVLVLEPARKESHGRYECGGDLDTMIS	330	
DB	272	ggnpppeefflylppgpegirssnlytlxd-----vrnalqdykoslidkksmla	323	
QY	331	LLSEPGQLLVNVYSDVAVSPAAP--EROEGSSLTITGCEASSODLEFQWLREETQGVLERG	389	
DB	324	sta-----ltvnyll-dslmnpsegtvqigdalprvscslasrnaevvmkdnlr--lrss	376	
QY	390	PVLOLHDLKREAGGGYRCVASVPSIPGLNFTQLVNVALFGPPMAFKRRKYWKKNVNLN	449	
DB	377	p--sfssihygdagynvcetalqevglkkresltllivegkp--qikmltktdpsglskt	432	
QY	450	LSCGASGHPRTTWMNNGTAS--EQDDOPQ--RVLSLMLVLTPELLEFGVECTASN	503	
DB	433	llchvegfpkpalqwlctlgsgsvvlnqdeespylngryys--kllspeernt-ltcaen	489	
QY	504	DLGKNTSLIFELVNLTLTPDSTNTTGLSTASPHTRANSTERRKLPEPESRGVIV	563	
DB	490	qlertvsnlvsalsi!--pe-----hdeadeidsendr-ekvndaklilv	530	
QY	564	AVIVCILLAVLGAVLFLV-KKKKLPCKRRSGQETLTPPSRKE	607	
DB	531	gltvgllllaalvagvywlymkksktskshvnxdklmgneenkkle	575	
RESULT	5			
W47088				
ID	W47088	standard; protein; 583 AA.		
XX	W47088;			
XX				
DT	20-JUL-1998	(first entry)		
XX				
DE	Activated leukocyte-cell adhesion molecule (ALCAM).			
XX				
KW	ALCAM; activated leukocyte-cell adhesion molecule; CD6 ligand;			
XX	antibody; human.			
OS	Homo sapiens.			
XX				
FH	Key	Location/Qualifiers		
FT	Peptide	1..27		
FT	Protein	/label= Sig-peptide		
FT		28..583		
FT		/label= Mat.protein		
FT	Domain	28..527		

Seq	Sequence	583 AA:
Query	Match	14.6%; Score 492.5; DB 19; Length 583;
	Best Local Similarity	24.8%; Pred. No. 9e-27;
	Matches	145; Conservative 121; Mismatches 256; Indels 63; Gaps 23;
QY	41 GSNALKCLSGSOGNLSHVDWFSVHKERKRLIFRVROGOGS----	96
DB	36 gdtliipclldvppq nlmfgkwyekpdsprfifisfstsktsvqddvpeykdtlnlse	94
QY	97 KQATLALQVTPDERIFELCO-GKRRSQEVRIOQLRVYKAEERNIQVNPGLGVNSKEP	155
DB	95 -lytlstisnatistekrficvmlytednfyfeapliavkfvfkqspkeivskalfi--eteql	151
QY	156 EEWAVCVGNNGPIPIQVILMYKNGRKLKEKNVNHQSSQYTES-SGLYTLDSILKAQLVK	214
DB	152 KKIIGCISDISEDpqnltwyrngkvlnhlegavvllfkkenndpytqltmtstleykltk	211
QY	215 EDKDAFYCELMRYRLPSGNHMKESREVIYVPEYPTREKWLLEY-EPVGLKLEGDRVEIRCL	273
DB	212 adigmpftcsvlyyypsgqkthlseqavfdiylpceqtlqvlppkxaikegdnltlkl	271
QY	274 ADGNPPH---FSIKQNPSTREAEETTTNDNGVLVLEPARKHSGRYECGLDIDTWIS	330
DB	272 gngnpppeefllylpqdgqeglrssntyltld-----vrrnatgdykcsllidkksmia	323
QY	331 ILSEPOELLVNVSVSVRSPAP-ERQEGSLTLTGCAESSODLEFQWLREITGOVLERG	389
DB	324 sta-----ltvhyll-dislnpsgevtirgldalpvscstisarnatvwmkdnlr--lrrs	376
QY	390 PVLQHLHLKREAGGGRCVASVPSIPGLNQTOLVNVALFGRPMWAFKRRKXWVENNVLN	449
DB	377 P--sfstshygdagnyvcetalqevgelkresltllivegkp-qikmktktldpsglskt	432
QY	450 LSCSEASGHRPTISMNVNGTAS---EQDQDQ---RVLSTIANVLVTEPLETGYECTASN	503
DB	433 ilchvegfpkpaigwtlctlgsgsvlnqcteespylmgtryys--kllispeenvt-ltctaan	489
QY	504 DLGKNTSLLEFELVNLTLTPDSNMTTGLSTSTASPHTRANSTSTERKLEPESRGVIV	563
DB	490 qlertvsnlsvalsist-----pe-----hdeadeisdenr-ekvndqakliiv	530
QY	564 AVIQCILVAVLGVAVLYFLY-KKGKLPCKRNSGKQELIILPPSRKSE	607
DB	531 givvgilllaalvagvyvlymkkksktskaskhnbkdlgmeeenkklie	575
RESULT	6	
ID	R97231	
AC	R97231 standard; Protein; 570 AA.	
AC	R97231;	
DT	08-OCT-1996 (first entry)	
DE	Stem cell marker HCAPro.2.	
KW	HCAPro.2.	
OS	Homo sapiens.	
PN	EP716146-A2.	
PD	12-JUN-1996.	
PF	02-FEB-1995; 95EP-0300661.	
PR	06-DEC-1994; 94US-0352323.	
XX	(SYST-) SYSTEMIX INC.	

QY 264 EGDREVEIRCLADGNPPPHFSISKONPSTREAEETTNDGVLVLEPARKHS 315
Db 1932 egdtltltcpiklaedlqadvmvswtkdsraldgltdhvdidsdgrkltlisqaalea 1991
QY 316 GREGGGLD-----IDTMSILSEPOELLVNVSDVRSPAPEROEGSSSLTTCFAES 369
Db 1992 gyltclalnragaslefkveilsp-----vldistndvqpyavngptlmrcavtg 2044
QY 370 SODLEFQWLR-----ETGVLEKGPVLQDLHLKREAGRCVAVSPISGLNRTOL 422
Db 2045 hpfpskwlknqkevdddenirivegqylqrlrtdshagkwscae--ndagvkelcm 2102
QY 423 VNVATGPPPMMAFERKRWKEMVNLSCFASGHRPRTISMNVNGTASQDDPQRYLS 482
Db 2103 v-ldvftppvsvskdnpikalgeltlfcnaagppqklwakaggslltdspdgatls 2161
QY 483 TLNVLTPELLETV-----ECTASNDLGKNTSILFELVNLTTLPDSNTTGLSTSPAS 539
Db 2162 kgarldipnlkktcdvqdytcgaalnaagltsea-----svsvdvlvpeindgid--msp 2213
QY 540 HFRANSTSTERKL-----PEPESR 558
Db 2214 rlpagqsltlqclagqkpyqmr 2236
RESULT 8
Y08402
ID Y08402 standard; Protein; 1380 AA.
XX
AC Y08402;
XX
DT 24-JUL-1999 (first entry)
XX
DE Drosophila sp. Robo2 extracellular domain protein.
XX
KW ROBO1; ROBO2; roundabout; nerve guidance; human; murine; cell function;
XX cell morphology; screening assay.
XX
OS Drosophila sp.
XX
PN W09920764-A1.
XX
PD 29-APR-1999.
XX
PF 20-OCT-1998; 98WO-US22164.
XX
PR 14-NOV-1997; 97US-0971172.
XX 20-OCT-1997; 97US-0062921.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman CS, Kidd T, Mitchell KJ, Tear G;
XX
DR WPI: 1999-312615/26.
XX N-PSDB: X57251.
XX
PT Robo polypeptides, a new immunoglobulin superfamily member
XX
PS Claim 1; Page 52-56; 80pp; English.
XX
CC This invention describes novel Robo (roundabout) polypeptides, involved
XX in nerve guidance which have been isolated from Drosophila sp.,
XX C. elegans, human and murine samples. The products of the invention can
XX be used to raise anti-Robo antibodies, which can be used to modulate cell
XX function or morphology. The Robo polynucleotides and fragments are useful
XX as probes and primers and for production of the Robo polypeptides. The
XX probes and primers are also useful in screening assays.
XX
SQ Sequence 1380 AA;

Best Local Similarity 23.8%; Pred. No. 2.6e-09;
Matches 119; Conservative 72; Mismatches 193; Indels 116; Gaps 27;
QY 137 EEPNIOVNPILGIVNKSKEPEVATCYGRNGYPLPQIWIYNGRPLLEEKNRVHIOSQY 196
Db 2 enprlienhpdltvpxkndp-ftncga-egnpelptlqikdgrlkdtdgshim-----1 55
QY 197 ESSGLYTLQSLILKAOLYKEKDAQFCELYRPLSGNHMESREVPVYPREKWLVEV 256
Db 56 paqglfllkvlhr-----resdagtlwceaknef-----gvartnalqavavide--fll 106
QY 257 EPVG-MIKEGDREVEIRCLAD-GNPPPHFSISKONPSTREAEETTNDGV 304
Db 107 epantlvaggevalmeagaprgsppqlsvrknq-----qtlvlgknkrlrldvggn 158
QY 305 LVLEPARKHSGRECOGLDLMTMSILSEPOELLVNVSDVRSP--AAPEROE--G 358
Db 159 latqearqsddgygc-----vknvnglresataflk-vhvrptllirgpnqtaavga 210
QY 359 SSLTINCE--AESQDL-----EFOWLREETGOV-LERGPVLQDLHLKREA 401
Db 211 ssavfgrtggdipdvltwrtasgmpmrkfwlmsagrvhvedrslkldvleed 270
QY 402 GCGYRCVAVSPISGLNRTOLVNVATGPPPMMAFERKRWKEMVNLSCFASGHRPRT 461
Db 271 mgeytcead-navggitatglltv--happkfvirpknqlveigdevlfecqanghprpt 327
QY 462 ISMNVNGTAS-----EQDQDQRYLSLNLVLTPELLET--GVECTASNDLGKNTSILF 515
Db 328 lywsevegsslllpgyrd-----grmexvlllpegrsvlsiarfareds9k----- 372
QY 516 LVNLTTLPDSNTTGLSTSPASPHRANSTSTERKLPEP-ESRGVIVAVIYCIYLAV 574
Db 373 vvtcnal-----navgsvsrtlv-----svdtcfelpppliegqpnqlpvksivv-- 420
QY 575 LGAVLYFLYKKRKLPCRSG 594
Db 421 -----lpcrtlg 427
RESULT 9
Y13564
ID Y13564 standard; Protein; 1381 AA.
XX
AC Y13564;
XX
DT 30-JUL-1999 (first entry)
XX
DE Drosophila Robo 2 polypeptide.
XX
KW Comm polypeptide; Robo polypeptide; commissureless; roundabout;
XX modulation; nerve cell function.
XX
OS Drosophila sp.
XX
PN W09925833-A1.
XX
PD 27-MAY-1999.
XX
PF 13-NOV-1998; 98WO-US24327.
XX
PR 14-NOV-1997; 97US-0065543.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Goodman C, Kidd T, Mitchell KJ, Russell C, Tear G;
XX
DR WPI: 1999-338008/28.
XX N-PSDB: X55768.
XX
PT Modulation of Robo-Comm polypeptide interactions
XX
PS Disclosure: Page 34-38; 56pp; English.


```

OY 321 OGDLDTMTSLSEPEOLLVNVSDVRVSPAAPERQ-----EGSSLTLTCEAES---SQ 371
DB 292 eyvnl---lgnrkveell-----vgafprdpelmsgglvngssvsvcskpsvypild 342
OY 372 DLEFQWLREETGOVLERGEVLOLHDLK-----REAGGRCYASV----- 411
DB 343 rlelletlkgct--lleniefledtdmkslenkslentfptletdgtkalvcqaklhiddm 400
OY 412 ---PSTPGLNKRLQLVNVAIFGPPMAFKERKRWKKNMVL-----NLSCEASGHRPTI 462
DB 401 efepkqrgstqcllyvna---p-----rdtvlvpspsllleegsvymtclsgfpapk1 452
OY 463 SMNVNGTASEODDOPRVLTSLNVLTPPELLETGVE-----CTASNDLGKNTSLILEL 516
DB 453 lvs-----rqlpngelqplsenatlcltstkmcdsgvylceglngqrstkevell1 504
OY 517 -----VNLTTLTLPDS-----NTTGLSTSTASPHT-----RANSTSTERK 551
DB 505 gvtpkdkikltafpsesvkegdvtlscctgnvpetwlllkkkeetgdtvlksldgaytlr 564
OY 552 -----LPEERSGVY-----IVAVYCIILVAVLGAV 578
DB 565 kaqlkdagvyecesknkysqrlsrltdvgrenknkdyfispellvlyfssallipalgm1 624
OY 579 LVFLYK---KGLPCRRSGKOEI 598
DB 625 lyfarkamkgsyislveaqskv 647

RESULT 11
R39682
ID R39682 standard; Protein: 736 AA.
XX
AC R39682;
XX
DE 23-DEC-1993 (first entry)
XX
DE VCAM/ICAM-2.
XX
KW Polymerase chain reaction; primer; PCR; amplify; VCAM-1; ICAM1;
KW vascular cell adhesion molecule; monoclonal antibody; chimera; 4B9;
KW VLA-4; expressing cells; intracellular adhesion molecule; VCAM-7D;
KW VCAM-6D; substitution; Ig superfamily; homology.
XX
OS Synthetic.
XX
FH Key
FH Peptide 1..24
FT /note="Signal peptide"
FT Protein 25..736
FT /note="Mature protein"
XX
XX WO9314220-A.
XX
XX 22-JUL-1993.
XX
XX 12-JAN-1993; 93WO-US00031.
XX
XX 13-JAN-1992; 92US-0821712.
XX
XX (BIOJ ) BIOGEN INC.
XX
XX Benjamin CD, Osborn L;
XX
XX WPI; 1993-243235/30.
XX
XX N-PSDB; Q43588.
XX
XX Monoclonal antibody to epitope on the fourth Ig-like domain of
XX VCAM-7D - for treating inflammation or disease associated with
XX leukocyte binding to endothelium e.g. post-reperfusion injury,
XX etc.
XX

```

```

PS Disclosure; Page 56-59; 108bp; English.
XX
CC This sequence represents a chimeric vascular cell adhesion molecule
CC (VCAM)/ICAM polypeptide which was used to determine regions of VCAM-1
CC involved in recognition of VLA-4-expressing cells. Recombinant genes
CC encoding VCAM/ICAM (intracellular adhesion molecule) chimeras were
CC produced by existing portions of the VCAM-1 molecule and VCAM-6D cDNAs and
CC replacing them with analogous regions of ICAM-1 using unique
CC restriction endonuclease sites within the VCAM-1 cDNA, and PCR
CC generated fragments of ICAM-1 (see also Q43578-84). Substitution
CC of the excised VCAM-1 regions was done to eliminate as much as
CC possible structural distortions that would accompany deletion of one
CC or more domains of the VCAM-1 molecule. ICAM-1 was chosen as a donor
CC because it is also a member of the Ig superfamily and has the highest
CC degree of amino acid homology to VCAM-1 among superfamily members.
CC The N-terminal three domain of VCAM-1 support adhesion to VLA-4, and
CC VCAM/ICAM-1 and VCAM/ICAM-2 in which most of the first domain has been
CC replaced by ICAM-1, do not bind to the anti-VCAM-1 monoclonal antibody
CC (Mab), 4B9. Therefore the epitope which binds to 4B9 must be situated
CC in domain 1.
XX
SQ Sequence 736 AA:
XX
OY 33 PELVEV-----EVGSLTLKCGLSQSGNLSHVDFSVHKEKRTLIFRVGGQSGSEP 85
DB 112 pervalaplpwpgvgnklltrc---qveggapran-----ltvlllrgelkmlksqef 161
OY 86 GEYEQRLSLQDRATLALQVTPQDERIFLCQKR-----PRQERYTOLRVYKAP 137
DB 162 ledadrtstetkslevlftprviedlqvlvcraklhiddmsvtrvqavkclvaylspk 221
OY 138 EPNIOVNPGLGIPVNSKEPEEVATCGVRNGYPIQOVLIWYK---NGRPLKEKNVHIOSSQ 194
DB 222 ntvisvnp---skllqegsgvmtcscseqlpapeiftfskklidng-----nlqhlsgna 271
OY 195 T-----VSSGLYTLQSLTILKAOLVKEKDQAFYCELNRLPSGNHMKESREVTVPVFY 247
DB 272 tlliammedsglylcycey---nll-----gynrke-----veliv 305
OY 248 PTEKWLVEVPVGM---KEDDRVEIRCLADGNPPPHSISKQNPSTREAEETTNDNGVL 305
DB 306 qekprfveaispripriaqldgsvmltcsymgscpsfswtqldspisgkvrsegtstl 365
OY 306 VLEPARKREHSGRYEC-----OGDLDTMTSLSEPEOLLVNVSDVRVSPAAPERQ--- 356
DB 366 tispvstenehsylctvtgchkklekgyivelsfpfd-----pelemsg 410
OY 357 ---EGSSLTLTCEAES---SQDLEFQWLREETGOVLERGEVLOLHDLK----- 398
DB 411 glvngssvsvcskpsvypildrlelletlkgct--lleniefledtdmkslenkslentf 468
OY 399 ---REAGGRCYASV-----PSTPGLNKRLQLVNVAIFGPPMAFKERKRWKKNMVL 447
DB 469 priedtqkalvcqaklhiddmetefepkqrgstqcllyvna---p-----rdtvlvpspsl 520
OY 448 L-----NLSCEASGHRPTISMNWNGTASEODDOPRVLTSLNVLTPPELLETGVE--- 498
DB 521 leegsvymtclsgfpapkllvs-----rqlpngelqplsenatlcltstkmcdsg 572
OY 499 ---CTASNDLGKNTSLILEL-----VNLTTLTLPDS-----NTTGLSTSTASPHT--- 541
DB 573 vylceglngqngarstkevelllqvtpkdkikltafpsesvkegdvtlscctgnvpetwll 632
OY 542 ---RANSTSTERK-----LPEERSGVY----- 561
DB 633 kkaeetdvlksldgaytlrkaqlkdagvyecesknkysqrlsrltdvgrenknkdyf 692
OY 562 ---IVAVYCIILVAVLGAVLVFLYK---KGLPCRRSGKOEI 598

```

Db 693 spellvlyfssallipaligmllyfarKamkysylveadsksv 736

RESULT 12

ID Y53666 standard; Protein: 4412 AA.

AC Y53666;

XX 22-FEB-2000 (first entry)

DE Sequence g1/1017427/emb/CAA62189 from an alignment with protein 608.

XX Mechanical stress; gene therapy; protein 608; osteoporosis; bone density;

KM bone development; g1/1017427/emb/CAA62189.

XX Unidentified.

OS

XX WO960164-A1.

PN

XX 25-NOV-1999.

PD

XX 14-MAY-1999; 99WO-US11066.

PF

XX 15-MAY-1998; 98US-0085673.

PR

XX (OUAR-) QUARK BIOTECH INC.

PA

XX Elinat P, Mor O, Skalter R, Feinstein E, Faerman A;

PI

XX WPI; 2000-053304/04.

DR

XX Identification of stress induced genes for determining risk and

XX preventing, treating or controlling osteoporosis -

PT

XX Claim 32; Fig 6A-R; 308pp; English.

PS

XX The present sequence is obtained from a clustal X alignment with

XX protein 608. Protein 608 was identified using the method of the invention

CC after subjecting rat osteoblasts to mechanical stress. Expression of the

CC 608 gene was found to be upregulated by about 3-fold in cells subjected

CC to mechanical strain. The specification describes a method for the

CC identification of genes responsive to a specific mechanical stress. The

CC method comprises applying the mechanical stress to an organism (tissue

CC or cells comprising bone cells), isolating the specific cellular

CC fractions and extracting mRNA from them, and differentially analysing the

CC mRNA in comparison with control samples. The method is used to identify

CC genes whose expression is responsive to a specific stress. The identified

CC genes are employed in determining risk associated with a physiological or

CC disease state. The risk determination methods are used for testing a

CC medication for gene therapy. These medications, or genes identified by

CC the method of the invention, are used for treating, preventing or

CC controlling a physiological or disease state (especially osteoporosis or

CC bone density or other factors causing or contributing to osteoporosis or

CC its symptoms or other conditions involved in mechanical stress or its

CC lack. The methods can also be used for advancing research or studies in

CC bone development.

CC

XX

SQ Sequence 4412 AA;

Query Match 7.0%; Score 236; DB 21; Length 4412;

Best Local Similarity 21.6%; Pred. No. 1.8e-07;

Matches 117; Conservative 88; Mismatches 231; Indels 106; Gaps 24;

QY 34 ELVEVEVSTALLKCGLSQSGNLSHWDFSVHKEKRTLLFRVQGGQSGSEPEYQURS 93

DB 1000 epleaavgdsvislqcyagvape--llvswy-----kgdklrp-eyrty 1042

QY 94 LQDRGATLALTYVTPQDERFLQ-----GRRRQSEYRQLVYK---APEPNQVNP 145

DB 1043 flnnvalvlnkynlndsgyctkaenslgtasskvtfrlqerqlpfpstfargldie-qt 1101

QY 146 LGIPVNSKEPEEVAACVGNHGYPIPOVIWYKNGRLKEEKNRHIOSSQTFESSGLYTLQ 205

DB 1102 vglpv-----tlcrlngsapi-gvccyrdgvalrrhen---lqsfvnvatlkilq 1150

QY 206 SILKAOLVEDKDAOFYCELNTRLDPSGNHMKESREVTVPVPEYKWLVEVPM-LKE 264

DB 1151 tdl-----shsgyscsasn--plgtasssar--ltarepkkspfldikpsidvia 1198

QY 265 GDRVEIRCLADGNPPHFSISKONPSTREAREETW-----NDNGVLVLPARKHSGREY 320

DB 1199 gesadtechnvgaqpmritwsdkmkeiirpagnyltlcvgnrphrlk-vgkgdsgytc 1257

QY 321 Q-----GLDLDTMISLSEPOELLVNVSDVPSAPABERQSSLTLTCEASSQDEF 375

DB 1258 gqndvkgkmcasqslsvkepk-----fkkleaskva---kgesigtleckisgspelkv 1310

QY 376 QMLREETG-----QVLERGPVLQHLDKREAGGGYRCVASYPSIFGINPQL-VNV 425

DB 1311 swfrndselheswkymfsfinsvallltlnaasaeedgyiceah---ngygdascstal 1366

QY 426 AIFGPPWMAFKRKRYVVENMYLNLSCENSGHPRRTISMVNVNGTASBEDQOPRYLSTLN 485

DB 1367 tvkappvftqkpsvgaalkgsdvlqcelstgfpfevww-----vkdrkqyrnsk 1417

QY 486 VLVTPELLEETGY-----ECTASNDLGKNT---SILFLELVNLTLPDSWTF 529

DB 1418 flttskhrfdtnlhlndeasdvgyehckatnevgsdltscsvkfkfpprfvkkisdctsl 1477

QY 530 TG 531

DB 1478 ig 1479

RESULT 13

ID Y88565 standard; Protein: 848 AA.

AC Y88565;

XX 07-AUG-2000 (first entry)

DE Human NCAM 140kd isoform precursor amino acid sequence.

XX

XX NCAM; neural cell adhesion molecule; Ig1; immunoglobulin domain 1;

KW neurite outgrowth promoter; proliferation; protein kinase C; sclerosis;

KW impaired myelination; stroke; Parkinson's disease; memory; schizophrenia;

KW Alzheimer's disease; diabetes mellitus; circadian clock; nephrosis;

XX treatment; prosthetic nerve guide; treatment; nervous system.

OS

XX Homo sapiens.

PN WO200018801-A2.

XX WO200018801-A2.

XX 06-APR-2000.

PD

XX 23-SEP-1999; 99WO-DK00500.

PF

XX 29-SEP-1998; 98DK-0001232.

PR 29-APR-1999; 99DK-0000592.

XX

XX (RONN/) RONN L C B.

PA (BOCK/) BOCK E.

PA (HOLM/) HOLM A.

PA (OLSE/) OLSEN M.

PA (OSTE/) OSTERGAARD S.

PA (JENS/) JENSEN P H.

PA (POUL/) POULSEN F M.

PA (SORO/) SOROKA V.

PA (RALE/) RALEYS I.

PA (BERE/) BEREZIN V.

PI Ronn LCB, Bock E, Holm A, Olsen M, OSTERGAARD S, Jensen PH;

PI Poulsen FM, Soroka V, Ralets I, Berezin V;
 XX WPI; 2000-293111/25.
 XX
 XX Compositions that bind neural cell adhesion molecules useful for
 PT treating disorders of the nervous system and muscles e.g. Alzheimer's
 PT and Parkinson's diseases -
 XX
 XX Disclosure: Fig 17; 119pp; English.
 XX
 XX This sequence represents the human neural cell adhesion molecule (NCAM)
 CC amino acid sequence. NCAM is found in three forms, two of which are
 CC transmembrane forms, while the third is attached via a lipid anchor to
 CC the cell membrane. All three NCAM forms have an extracellular structure
 CC consisting of five immunoglobulin domains (Ig domains). The Ig domains are
 CC numbered 1 to 5 from the N-terminal. The invention relates to a compound
 CC containing a peptide which binds to the NCAM Ig1 domain. The compound
 CC binds to NCAM-Ig1/Ig2 domains, and is capable of stimulating or promoting
 CC neurite outgrowth from NCAM presenting cells, and is also capable of
 CC promoting the proliferation of NCAM presenting cells. The compound may be
 CC used in the treatment of normal, degenerated or damaged NCAM presenting
 CC cells. The compound may in particular be used to treat diseases of the
 CC central and peripheral nervous systems such as post operative nerve
 CC damage, traumatic nerve damage, impaired myelination of nerve fibres,
 CC conditions resulting from a stroke, Parkinson's disease, Alzheimer's
 CC disease, dementia, sclerosis, nerve degeneration associated with
 CC diabetes mellitus, disorders affecting the circadian clock or
 CC neuro-muscular transmission and schizophrenia. Conditions affecting the
 CC muscles may also be treated with the compound, such as conditions
 CC associated with impaired function of neuromuscular connections
 CC (e.g. genetic or traumatic shock or traumatic atrophic muscle disorders).
 CC Conditions of the gonads, pancreas (e.g. diabetes mellitus types I and
 CC II), kidney (e.g. nephrosis), heart, liver and bowel may also be treated
 CC using the compound. The compound is used in a prosthetic nerve guide, and
 CC also to stimulate the ability to learn, and to stimulate the memory of a
 CC subject.
 CC
 XX
 XX Sequence 848 AA:
 SQ

Query Match 7.0%; Score 234; DB 21; Length 848;
 Best Local Similarity 21.1%; Pred. No. 2.4e-08;
 Matches 126; Conservative 86; Mismatches 222; Indels 164; Gaps 31;

QY 33 PELVEVEVSTALLKGLSGSGNLHVD--WFSVHEKRTLIFRVQGGOGSEPEYEQ 90
 DB 26 psggeisvsgskfflc---qvagdakdkdswfsngklt-----pnqg 67
 QY 91 RLSTL---QDRGATLALQVTPQDERIFLC--OGKRPRSQEYRQLRVYKAPERNIQVP 145
 DB 68 rlsvvvndssstltynanidaglykcvvgedgseseatvnrklfg----klmfkn 122
 QY 146 LGIPVNSKEEP--VATCVGNGGYPIPOVITWKNRPLKEKN--RVHQSQVTESSGLT 203
 DB 123 aptpgelregedavldcvvssip-pltllkhkgivlilkdrifvlisnnylqitg--- 178
 QY 204 LGSILKAQLVKEKXDAOFCELVNLRPSGNHKKESREYVVPVPTPEKYWLEVEYEMLK 263
 DB 179 -----ikktdeglytce-----grllarge-----infldiqylvnppliqgr 217
 QY 264 E-----GDRVEIRCLADGNPPHFSISKONPSTRAEET-----TNDGVLVLEPA 310
 DB 218 gnlvntanlqsgvltvcdagdfprrtmwtckdgediegedekyifsdssqlltkkv 277
 QY 311 KRHSRRYECG---LDLDTMISLSEPEQLLVNVSDVVSAPAEREGSSLLT 364
 DB 278 dkndaeeyiclaenkagqdatlhkvfakp---itye-----pqtmelcee---qvlit 328
 QY 365 CEAVESQDLEFQW-----LREET--GQVLEKRPV-----LQLDLKRKAGGRCVCA 410
 DB 329 ceasgdpipsticwtstnmsiseekldghmvvsharvstllkslygtdeageyictas 388
 QY 411 -----VPSIPGLNRTQLVNVAIFGPPMMAFKERKRVWVKENNVLNLSCA 454

DB 389 ntlgqdsqsmylexyaprklyg-----pvavy-----tw---egqvnltcev 428
 QY 455 SGHPRPTISWNVGTASEQDQDQPVRLSTLVNVTP--ELLEF-----GVECTASN 503
 DB 429 faypsatlswfirgqll-----psnynsklynrpsasylevtpdsendfynynctavn 483
 QY 504 DLGKNTSILFLVLVNI--TTLTPDSNTTGTGSTASPHTRANSTGERKLPEPESRGV 561
 DB 484 rlyges-----lefilvgadtppsp-----slgdvepy-----stlaqvgtdepatgyv 528

RESULT 14
 R08117
 ID R08117 standard; protein; 647 AA.
 XX
 XX R08117;
 XX
 XX 27-FEB-1991 (first entry)
 DE Vascular cell adhesion molecule 1 from pCDM8 clone 41.
 XX
 XX Endothelial cell-leucocyte adhesion molecule 1; ELAM1;
 KW vascular cell adhesion molecule; VCAM1; antibodies;
 KM molecule involved in leucocyte adhesion; M1A; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN W09013300-A.
 PD 15-NOV-1990.
 XX
 PD 27-APR-1990; 90WO-0502357.
 PF 18-DEC-1989; 89US-0452675.
 PR 28-APR-1989; 89US-0345151.
 PR 01-JUN-1989; 89US-0359516.
 XX
 PA (BIOG-) BIOGEN INC.
 XX
 PI Hession C, Lobb RR, Goeltz SE, Born L, Benjamin CD;
 PI Rosa MD;
 DR N-PSDB; Q06687.
 DR WPI; 1990-361248/48.
 XX
 PT Endothelial cell adhesion mols. - M1As and DNA encoding them and
 PT inhibition-detection of binding of leukocytes to endothelial
 PT cells
 PS
 XX
 XX Disclosure: Fig 3 (A-D); 136pp; English.
 CC The VCAM1 polypeptide possesses a hydrophobic N-terminal
 CC characteristic of a signal sequence. It is predicted that the
 CC N-terminal amino acid of the mature protein will be phenylalanine 25.
 CC The extracellular domain of the polypeptide is ca. 606 amino acids
 CC including the signal sequence and is followed by a hydrophobic
 CC transmembrane region of 22 amino acids. The protein possesses a
 CC short charged cytoplasmic tail of 19 amino acids. The protein
 CC contains six potential N-glycosylation sites.
 CC Comparison of VCAM1 and VCAM1b (Q06688) revealed that they are
 CC virtually identical except for one significant difference: VCAM1b
 CC contains an insertion of 276 nucleotides near the middle of the
 CC coding region. These nucleotides encode 92 additional amino acids
 CC which form an extra domain of 84 amino acids situated between the
 CC end of VCAM1 domain 3 and the beginning of VCAM1 domain 4. This
 CC domain is designated domain 3B.
 CC Cells expressing ELAM can be used to identify mols. which inhibit
 CC binding of leucocytes to endothelial cells, and such binding can
 CC be inhibited (i.e. inflammation can be treated) using an ELAM
 CC (or fragment), antibodies which recognise M1A, ELAM ligands or
 CC their fragments, carbohydrates which bind to ELAM and antibodies
 CC which recognise ELAM.

CC See also Q06686-91.
XX
SO Sequence 647 AA;

Query Match 6.9%; Score 233; DB 11; Length 647;
Best Local Similarity 21.9%; Pred. No. 2e-08;
Matches 134; Conservative 89; Mismatches 234; Indels 156; Gaps 27;

```

QY 99 ATLALTOYTPODERIFLCOGK-RPRSGYRIOLRYKAPPEPNIOVN-PL-GIPVNSKE 154
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 78 stllmpvysfgnehytltaccesrkleqiyvelysfpkdpelhsplaeagpflvk- 136
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 155 PEEVATCGVRNGYRPIQY-IWYKNGRPLKEKNRYHIOSQTVSSGCL-YTLOSILKAQL 212
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 137 -----cavadypfdarldelldlkghdhlmsqefledadrsletkslevftfpvl---- 186
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 213 VKEDDAQFYCELNRLPLSGNHMKESREVT--VPYFYTEKVMLEVEPVGMLEKGDREVEI 270
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 187 --edlgkylvcraklhidemsvprvqavkelyisphkntvispstkldggsvtm 244
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 271 RCLADGNPPHFSISKONPSTREAEETTNDNGVLVLEPARKHSGRYECGLDLDTMIS 330
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 245 tcssegllpepelfwskk---ldngnlqhsnatltllammedsglyvcegvnl---ig 298
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 331 LUSEPOELLVNVSDVRVSPAAPERQ-----EGSSLTLCFAES---SODLEFQWLREE 381
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 299 knrkevell-----vqefrpdrelemsgglvngssvtscvpsvypldrlelellkge 352
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 382 TGOVLEGRGVNLDLHDK-----REAGGGRVAVSV-----PSIPGLN 418
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 353 l--llenlefledtmkslenkslemftiptledqkaivqaklhidemetepqrqsc 410
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 419 RQOLVNAVAFGPPMAFKERKYVKNENYL-----NISCEASGHPRPISNNVGTASE 472
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 411 gelynvva-----p-----rdctvlvpspslleegssvmtclisqgfpakklws----- 455
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 473 QODDOPRVSTLNVLTVPPLLELGEV-----CTASNDLGKSTLIFEL-----VNLT 520
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 456 -rqlprgelqplsenatcltistkmedsgvylceqngqgrtkewelllyqtpdklkt 514
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 521 TLTTPDS-----NTTGTGSTASPHRT-----RANSTSTERK-----LP 553
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 515 afpsesevkgdvtlascctgnvpetvlllkkkaetgdvllksldgyltrkaqlkdagvy 574
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 554 EPPSRGVY-----IYAVIVCIIVLVGLVFLVK---K 585
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 575 ecesknkvsgqlrslldvgreenkdyfspellvlyfassllipalgmilyfarkamk 634
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 586 GKLPCCRSGKQEI 598
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 635 gysylveagkskv 647
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

```

RESULT 15
ID R38549
R38549 standard; Protein: 647 AA.

AC R38549;
DT 23-DEC-1993 (first entry)

DE VCAM-6D.

XX Polymerase chain reaction; primer: PCR; amplify: VCAM-1; ICAM1;
XX vascular cell adhesion molecule; monoclonal antibody; chimera: 4B9;
KM VLA-4; expressing cells; intracellular adhesion molecule; VCAM-7D;
KM VCAM-6D; substitution; Ig superfamily; homology.

XX Synthetic.

XX Key Location/Qualifiers
FT Peptide 1..24

FT /note= "Signal peptide"
FT Protein 25..647
FT /note= "Mature protein"

XX MO9314220-A.

XX 22-JUL-1993.

XX 12-JAN-1993; 93WO-US00031.

XX 13-JAN-1992; 92US-0821712.

XX (BIOG) BIOGEN INC.

XX Benjamin CD, Osborn L;

XX WPI; 1993-243235/30.

XX DR N-PSDB; Q43586.

PT Monoclonal antibody to epitope on the fourth Ig-like domain of
PT VCAM-7D - for treating inflammation or disease associated with
PT leukocyte binding to endothelium e.g. post-reperfusion injury,
etc.

PS Disclosure; Page 43-46; 108pp; English.

XX This sequence represents vascular cell adhesion molecule (VCAM-6D).
CC This sequence was used in the production of chimeric VCAM/ICAM
CC polypeptides which were used to determine regions of VCAM-1 involved
CC in recognition of VLA-4-expressing cells. Recombinant genes encoding
CC by VCAM/ICAM (intracellular adhesion molecule) chimeras were produced by
CC existing portions of the VCAM-7D and VCAM-6D cDNAs and replacing them
CC with analogous regions of ICAM-1 using unique restriction endonuclease
CC sites within the VCAM-1 cDNA, and PCR generated fragments of ICAM-1
CC (see also Q43578-84). Substitution of the excised VCAM-1 regions was
CC done to eliminate as much as possible structural distortions that
CC would accompany deletion of one or more domains of the VCAM-1
CC molecule. ICAM-1 was chosen as a donor because it is also a member
CC of the Ig superfamily and has the highest degree of amino acid
CC homology to VCAM-1 among superfamily members. The N-terminal three
CC domain of VCAM-1 support adhesion to VLA-4, and VCAM/ICAM-1 and
CC VCAM/ICAM-1, do not bind to the first domain has been replaced by
CC 4B9. Therefore the epitope which binds to 4B9 must be situated in
CC domain 1.

XX Sequence 647 AA;

Query Match 6.9%; Score 233; DB 14; Length 647;
Best Local Similarity 21.9%; Pred. No. 2e-08;
Matches 134; Conservative 89; Mismatches 234; Indels 156; Gaps 27;

```

QY 99 ATLALTOYTPODERIFLCOGK-RPRSGYRIOLRYKAPPEPNIOVN-PL-GIPVNSKE 154
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 78 stllmpvysfgnehytltaccesrkleqiyvelysfpkdpelhsplaeagpflvk- 136
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 155 PEEVATCGVRNGYRPIQY-IWYKNGRPLKEKNRYHIOSQTVSSGCL-YTLOSILKAQL 212
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 137 -----cavadypfdarldelldlkghdhlmsqefledadrsletkslevftfpvl---- 186
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 213 VKEDDAQFYCELNRLPLSGNHMKESREVT--VPYFYTEKVMLEVEPVGMLEKGDREVEI 270
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 187 --edlgkylvcraklhidemsvprvqavkelyisphkntvispstkldggsvtm 244
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 271 RCLADGNPPHFSISKONPSTREAEETTNDNGVLVLEPARKHSGRYECGLDLDTMIS 330
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 245 tcssegllpepelfwskk---ldngnlqhsnatltllammedsglyvcegvnl---ig 298
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 331 LUSEPOELLVNVSDVRVSPAAPERQ-----EGSSLTLCFAES---SODLEFQWLREE 381
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 299 knrkevell-----vqefrpdrelemsgglvngssvtscvpsvypldrlelellkge 352
   ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

```

```

0Y 382 TCOVLERGVLTOLHLC-----RACGGRYCVASV-----PSLPGLN 418
    : : : : :
Db 353 t-lleniefedctadmslenksjlemftfirtedtykalyocaklhdidnefeprqrst 410
    : : : : :
0Y 419 KQULYNVAIFGPPWMAFERKRYVWKENVLT-----MLSCBAGHPPTISMVWNGTASE 472
    : : : : :
Db 411 qlylvuva-----p-----rdttvlvpspslleegssvmtctsqfpeapklivs----- 455
    : : : : :
0Y 473 ODOOPORVLTSLANLVYPELLEVE-----CRASNDLCKNLSIFEL-----VNTF 520
    : : : : :
Db 456 -rqjpnngelqjlsenaltltliscmedsgvyjceghnqagrsrkevelilqvtrpkdikt 514
    : : : : :
0Y 521 TLTPDS-----NTTGLSTASAPST-----RANSTSTERK-----LP 553
    : : : : :
Db 515 afpesvtekgdltvliscctgnvpetwlilkkkaetgdltvksidagyltirkaqlkdagvy 574
    : : : : :
0Y 554 EPESRGVV-----IVAVICILVLANVAVLYFLXK--K 585
    : : : : :
Db 575 ecesknkvysqglrtltdvgremndkyfepellvlyfaasliipalgmlylfarkamnk 634
    : : : : :
0Y 586 GKLPGRSGRKEI 598
    : : : : :
Db 635 gsyslveaqkskv 647

```

Search completed: March 23, 2001, 12:31:40
Job time: 63 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2001, 12:30:37 ; Search time 22.65 Seconds
(without alignments)
1936.593 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363

Sequence: 1 MGIPRLVCATFLAACCCCPR.....SSGDKRAPGDQGEYIDLRH 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 6790655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3327	98.9	646	2	cell surface glyco
2	863	25.7	584	2	s-glycerin precurs
3	647.5	19.3	628	2	Lutheran blood gro
4	606	18.0	588	2	B-CAM protein - hu
5	509.5	15.2	588	2	adhesion molecule
6	505.5	15.0	588	2	surface glycoprote
7	502.5	14.9	587	2	DM-GRASP precursor
8	492.5	14.6	583	2	alcam - human
9	457	13.6	523	2	neurotin - goldf
10	281	8.4	5175	2	hypothetical prote
11	281	8.4	5198	2	hemilectin precurs
12	270	8.0	764	2	irregular chiasm C
13	264.5	7.9	538	2	vascular cell adhe
14	259.5	7.7	1051	2	kinase-like protei
15	256	7.6	1088	1	neural cell adhesi
16	251.5	7.5	4331	2	perlecan precursor
17	247	7.3	1612	2	ductal protein - mo
18	246.5	7.3	725	2	neural cell adhesi
19	246.5	7.3	739	2	vascular cell adhe
20	246.5	7.3	1651	2	transmembrane rece
21	244.5	7.3	858	1	neural cell adhesi
22	244.5	7.3	3707	2	heparan sulfate pr
23	242.5	7.2	853	1	neural cell adhesi
24	237	7.0	725	2	neural cell adhesi
25	237	7.0	1092	1	neural cell adhesi
26	236	7.0	7962	2	elastic titin - hu
27	235	7.0	1091	1	neural cell adhesi
28	234	7.0	761	1	neural cell adhesi
29	233.5	6.9	1115	1	neural cell adhesi

ALIGNMENTS

Result	1	138049	cell surface glycoprotein MUC18 precursor - human
N:Alternate names:			melanoma-associated glycoprotein MUC18 precursor
C:Species:			Homo sapiens (man)
C:Date:			01-Mar-1996 #sequence, revision 01-Mar-1996 #text_change 21-Jul-2000
C:Accession:			138049; A34507
R:Refs:			C.; Kirsch, K.; Rothbacher, U.; Rietmuller, G.; Johnson, J.P.
Proc. Natl. Acad. Sci. U.S.A.			90, 8514-8518, 1993
A:Title:			Genomic organization of the melanoma-associated glycoprotein MUC18: implicat
A:Reference number:			138049; MUID:93391384
A:Accession:			138049
A>Status:			preliminary
A:Molecule type:			DNA
A:Residues:			1-646 <RES>
A:Cross-references:			EMBL:X68264; NID:9433891; PIDN:CAA48332.1; PID:9825693
R:Lehmann, J.M.; Rietmuller, G.; Johnson, J.P.			Proc. Natl. Acad. Sci. U.S.A. 86, 9891-9895, 1989
A:Title:			MUC18, a marker of tumor progression in human melanoma, shows sequence simil
A:Reference number:			A34507; MUID:90099368
A:Accession:			A34507
A:Molecule type:			MRNA
A:Residues:			1-71, 'SSVCARARARANGSTK', 91-120, 'LGRSTASSASTKLRM', 139-587, 'AAVALREAG
A:Cross-references:			GB:M29277; GB:M28882
A>Note:			this sequence has been corrected in 138049
C:Genetics:			
A:Gene:			GDB:MCAM; MUC18; CD146
A:Cross-references:			GDB:304548; OMIM:155735
A:Insertions:			23/1; 64/3; 134/1; 157/3; 187/1; 287/3; 342/1; 381/3; 429/1; 469/3;
C:Keywords:			glycoprotein; transmembrane protein
Query Match			98.9%; Score 3327; DB 2; Length 646;
Best Local Similarity			98.9%; Pred. No. 2, 6e-205;
Matches			639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Db	1	MGIPRLVCATFLAACCCCPRVAGVGEAEQAPAELEVEVGSSTALLKCGISQSGMLSHV	60
Qy	1	MGIPRLVCATFLAACCCCPRVAGVGEAEQAPAELEVEVGSSTALLKCGISQSGMLSHV	60
Db	61	DWFSVHKERTLLFRVVGOGGSEPGVEEGRSLDORGATLTOVTPDERIFLCQGR	120
Qy	61	DWFSVHKERTLLFRVVGOGGSEPGVEEGRSLDORGATLTOVTPDERIFLCQGR	120
Db	61	DWFSVHKERTLLFRVVGOGGSEPGVEEGRSLDORGATLTOVTPDERIFLCQGR	120
Qy	121	PRSGEIRIQLRVYKAPPEPNIQVNPGLIPVNSKEPEVATCVGRNGYPIPOVITWYKNGRP	180
Db	121	PRSGEIRIQLRVYKAPPEPNIQVNPGLIPVNSKEPEVATCVGRNGYPIPOVITWYKNGRP	180
Qy	181	LKEKNRVHVIQSSQYVSSGLYTLQSLTKAQLVKEKDKAQFCELYRILPSCNHMKESRE	240
Db	181	LKEKNRVHVIQSSQYVSSGLYTLQSLTKAQLVKEKDKAQFCELYRILPSCNHMKESRE	240

Db 296 SPSEPTLFR-----LQDEQEVLVNLEGNLTLEGVTRGSGTYGCRVEDYDAADVQLS 351
QY 334 EPOELLVNVSDVRSVPAAPERQEG-----SSLTLTCEAESSODLEFQWIREETGOV 385
Db 352 KTELEFRAVTLDPLELS-----EGKVLSTPLNNSAVVNCVHGILPTPALRMKTDST--P 402
QY 386 LERGPVLQDLHDKREAGGGRVCAVSPISPLGNRTQLVNVAIFGPPMAFERK-----V 440
Db 403 LQDQPMILSTSTFDSNGTYCEASLPTVPLVSRQNTFTLVGSGPELKTAEIEPKADGS 462
QY 441 WKENMVLNLSCEASGHPRTTSMNVNGTASEDDQDPOR---VLSTLVNLVTPPELLETGV 497
Db 463 W-REDEVTLLICSAHGHPDKLMSQLG--GSPAEPPIPGRCGWSSSLTKTALSRLDGI 520
QY 498 ECTASNDLQKNTSLFLELVNLFTLTPDSNTTGLSTSTASPHTRANSTERKLPEPES 557
Db 521 SCEASNPHGKRVHVFHGAVS-----PQTSQ 546
QY 558 RGVVIAVAVICILVAVLGVAVLYFLYKKGLPCRRSGKQETLPPSRKSELVEYKSDKL 617
Db 547 AGVAVMAVAVSVGLLLVAVFVYCVRRKG--PCCRQRRREGAPP---GEPGLSHSGSEQ 602
QY 618 PEEMGLIQQ--SSGDKRAPGDGCK 640
Db 603 PEQTGLIMGASGARGSGGFGDE 627

RESULT 4
137202
B-CAM protein - human
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000
C:Accession: I37202; S47272
R:Campbell, I.G.; Roulees, W.D.; Senger, G.; Trowsdale, J.; Garin-Chesa, P.; Rettig, W.J.
Cancer Res. 54, 5761-5765, 1994
A:Title: Molecular cloning of the B-CAM cell surface glycoprotein of epithelial cancers:
A:Reference number: I37202; MUID:95042297
A:Accession: I37202
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-588 <RES>
A:Cross-references: EMBL:X80026; NID:9535178; PIDN:CAAS6327.1; PID:9535179
C:Genetics:
A:Gene: B-CAM

Query Match 18.0%; Score 606; DB 2; Length 588;
Best Local Similarity 29.2%; Pred. No. 2.9e-31;
Matches 187; Conservative 96; Mismatches 247; Indels 110; Gaps 25;
QY 2 GLPR-LVCAFLLAACCCCPRVAGVGEAQPAPLVEVVGSTALIKGLSOGNLSH- 59
Db 12 GAFRLLLAVLLAA-----HPDAQAEVNLVSPPLVEVNRKGSVLLDC---TPGTGHHY 62
QY 60 -VDMF-----SVHKEKRLIFRVROGQSGSEPEYEOURLSLQDRGATLALQV 106
Db 63 MELEWFLTRSGARPRLASAEQSGELQVMTMDLGRSP---YLLDSGR---LVLAFA 115
QY 107 TQODEEIFLC--QGRKRQOEYRIQLRYKAPKEEPNIOVNPGLIVNKSKEPEVATCYGR 164
Db 116 QVGEDEDVYCVVRAAGAATAELARLNFAPKEATEVSPNKSTLLEVMEDSAOEIATCNSR 175
QY 165 NGCVPIQVWYKNGRPLK--EKKNRVHIQSQTV--ESSGLTYLQSLIAQLVKEKDKNOF 221
Db 176 NQNPAPKIIWYNGQRLVPEVMPNPEGVITSTVREASGLSLSTLVLPCKKDRDASF 235
QY 222 YCELNRLPLSGNHNK--ESSEVTVPVFVPTKY--WL--EVEPVGMKLGSDRYEINCLAD 276
Db 236 HCAAHYSLPEGHGRIDSPFTFLTLHYPTREHYQFWGSPSPAGVWRBDGVQLLCRGDG 295
QY 277 NPPPHSISKQNPSTREAEETTNDN--GVVLPEPARKHSGRYCCGGLDLDPMISLSLSE 334
Db 296 SPSEPTLFR-----LQDEQEVLVNLEGNLTLEGVTRGSGTYGCRVEDYDA----- 344

QY 335 POELLVNVSDVRSVPAAP-ERQEG-----SSLTLTCEAESSODLEFQWIREETGOV 385
Db 345 ADDVQLSKTLIDVRAVYLDPLLESEKVLSTPLNNSAVVNCVHGILPTPALRMKTDST--P 402
QY 386 LERGPVLQDLHDKREAGGGRVCAVSPISPLGNRTQLVNVAIFGPPMAFERK-----V 440
Db 403 LQDQPMILSTSTFDSNGTYCEASLPTVPLVSRQNTFTLVGSGPELKTAEIEPKADGS 462
QY 441 WKENMVLNLSCEASGHPRTTSMNVNGTASEDDQDPOR---VLSTLVNLVTPPELLETGV 497
Db 463 W-REDEVTLLICSAHGHPDKLMSQLG--GSPAEPPIPGRCGWSSSLTKTALSRLDGI 520
QY 498 ECTASNDLQKNTSLFLELVNLFTLTPDSNTTGLSTSTASPHTRANSTERKLPEPES 557
Db 521 SCEASNPHGKRVHVFHGAVS-----PQTSQ 546
QY 558 RGVVIAVAVICILVAVLGVAVLYFLYKKGLPCRRSGKQETLPPSRKSELVEYKSDKL 617
Db 547 AGVAVMAVAVSVGLLLVAVFVYCVRRKG--PCCRQRRREGAPP---GEPGLSHSGSEQ 602

RESULT 5
JH0506
adhesion molecule SC1 precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 21-Jul-2000
C:Accession: JH0506; PS0270
R:Tanaka, H.; Matsui, T.; Agata, A.; Tomura, M.; Kubota, I.; McFarland, K.C.; Kohn, B.
Neuron 7, 535-545, 1991
A:Title: Molecular cloning and expression of a novel adhesion molecule, SC1.
A:Reference number: JH0506; MUID:92030150
A:Accession: JH0506
A:Residues: 1-588 <TAN>
A:Cross-references: GB:S63276; NID:9238000; PIDN:AMB20170.1; PID:9238001
A:Experimental source: embryo
A:Accession: PS0270
A:Molecule type: protein
A:Residues: 34-48 <TAN1>
C:Comment: This protein is uniquely and transiently expressed on spinal cord motoneur
C:Keywords: glycoprotein; transmembrane protein
F:1-33/Domain: signal sequence #status predicted <SIG>
F:34-588/Product: adhesion molecule SC1 #status predicted <ADH>
F:500-523/Domain: transmembrane #status predicted <TRA>
F:101,173,199,271,312,366,462,485,504/Binding site: carbohydrate (Asn) (covalent) #st

Query Match 15.2%; Score 509.5; DB 2; Length 588;
Best Local Similarity 26.4%; Pred. No. 4.3e-25;
Matches 166; Conservative 117; Mismatches 254; Indels 91; Gaps 30;
QY 6 LVCAFLLAACCCCPRVAGVGEAQPAPLVEVVGSTALIKGLSOGNLSHVDM-FS 64
Db 18 LLC--LLLAALCMPPALGL-----YYNNAVYQGITIMPCKRLVPEPD-LMFGKKYIE 65
QY 65 VHKERKTLI-FR--VRQGGQSGSEPEYEOURLSLQDRGATLALQVTPQDERIFLCQ--CK 119
Db 66 MPNSPVEIARFSTKKNVQYDVPDKRDLSE--NYTSLIKNARISDEKRRVCMLVTE 124
QY 120 RPRQOEYRIQLRYKAPKEEPNII--QVNPGLIPVNSKREPEVAVCVGNGVPTQVQVWYKN 177
Db 125 DQVSEEPVY-VKVRQPSQPEILHQAFLL---ETELKMLGECVNDSPPEGNVWYKN 179
QY 178 GRPLK--EKKNRVHIQSQTVESGLTYLQSLIAQLVKEKDKDAQFCELYNRLPSGNHM 235
Db 180 GRVLPQPEEVVYVILRKVEN--RSTGLFTMTSSIQYMPKTKDAKNAFTCIYTYHGPSGQRT 238
QY 236 KESKEVYVAVFVPEKAWLEY--EPVGMKLGSDRYEINCLADNPPPH---FSISKQNPST 291
Db 239 IQSEPVVEDVHYPTKEITIVLSQSSPTIKGDVNTLKSGNGNPPPOEFLFYI----- 291
QY 292 REAEETTNDNGVVLPEPARKHSGRYCCGGLDLDPMISLSPOELLVNVSDVRSVPA 351

Db 292 -PGETGIRSSDYVMTDVRNRATGEGYKCSLIDKSM-----DATITVHYL-DLQLPFS 344
Qy 352 AP-EROGSSSLITLCEASSQDLFEQWLREFTGVLEKRPVLQDLKREAGGRCVNS 410
Db 345 GEVTKQIGALPVSCTISSSRNATVFWIKDNTR--MKTSP--SFSSLOYODAGNICETT 400
Qy 411 VPSIPLNRTOLVNVALFPGPMMAFKERKVVYKENVNLNLSCEASGHPRTISMVNGTA 470
Db 401 LOEVEGLKRRKTKLIVEGRP--QIKMTKKTNTNKKMKTIVCHVEGFPKPAQVITGSG 458
Qy 471 S--EODODPQVRLSTLN--VLVTPLELLETGVECTASNDLGKNTSILFLELVNLTLLPDS 526
Db 459 SLINKTEETKYVNGKFSKRIIAPENVT-LTICIAENEL-----E 497
Qy 527 NTTGLSTAS-----PHTRANSTSTERKLPPEPSRGVIVAVIYVILVAVLAV 580
Db 498 RIVTSLNVAISAIPIEYDEPEDR--NDONSEK---VNDQAKLIVGIVGLLVALVAVY 552
Qy 581 FLY-KKGLPCRRSGKEITLPPSRKSE 607
Db 553 WLTVKSKTASKAVHDKDLGNIEENKLE 580

RESULT 6
A45254
surface glycoprotein BEN precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 21-Jul-2000
C:Accession: A45254, S19202
R:Pourquie, O.; Corbel, C.; Le Caer, J.P.; Rossier, J.; Le Douarin, N.M.
Proc. Natl. Acad. Sci. U.S.A. 89, 5261-5265, 1992
A:title: BEN, a surface glycoprotein of the immunoglobulin superfamily, is expressed in
A:reference number: A45254; MUID:92302224
A:Accession: A45254
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-588 <POU>
A:Cross-references: EMBL:X64301; NID:g63087; PIDN:CAM45579.1; PID:g63088
C:Keywords: glycoprotein

Query Match 15.0%; Score 505.5; DB 2; Length 588;
Best Local Similarity 26.4%; Pred. No. 7.7e-25;
Matches 166; Conservative 116; Mismatches 255; Indels 91; Gaps 30;

Qy 6 LVCAFLAACCCCPRVAGVGEAEQAPAEVVEVGSTALLKCGLSQSGNLSHYDW-FS 64
Db 18 LLC--LLLAALCMPRALG-----YVNAVYGDITMPCRLVEPDG-LMEGKMKYE 65
Qy 65 VHKERTLI-FR--VRQGGGSEPEYEORLSLQDRGATLALQVTPQDERIFLCO--GK 119
Db 66 MPNGSPVFIAPFRSTKKNVQYDDVYDKRLSLSE--NYTSLIKNARIHDEKRFVCMVTE 124
Qy 120 RPRSOEYRIQLVYKAPPEPNI--QVNPILGIPVNSKEPEEVATCGVGRNGYPIPOVIWYKN 177
Db 125 DVVSEEPYV-VKVFQPSPELHQADFL----ETEKLMKLGECVVRDSYFPEGNTVWYKN 179
Qy 178 GRPLK--EENRNVIHQSSQVYESSGLYTLQSLKAQLYKDKDAQFCELYNRLPSGNHM 235
Db 180 GRVLQPEVEEVVNLNRKEN--RSTGLFTMTSLOYMPTKEDANNAKFTCIYTHGSPSQKT 238
Qy 236 KESREYTVAVFPTEKVMLEV--EPVGMLEKEDGRVETICLADGNPPH--FSISKQNPST 291
Db 239 IQSEPVFVDVHYPTREKTIIVLSSQSTIKRGDNVTLKCSGNGNPPQEFLEYI----- 291
Qy 292 REAEETTNQNGVLVLEPARKHSGRYECQGLDITMTLSLSEPOLLVNVYSDVRSVA 351
Db 292 -PGETGIRSSDYVMTDVRNRATGEGYKCSLIDKSM-----DTTITVHYL-DLQLPFS 344
Qy 352 AP-EROGSSSLITLCEASSQDLFEQWLREFTGVLEKRPVLQDLKREAGGRCVNS 410
Db 345 GEVTKQIGALPVSCTISSSRNATVFWIKDNTR--MKTSP--SFSSLOYODAGNICETT 400

Qy 411 VPSIPLNRTOLVNVALFPGPMMAFKERKVVYKENVNLNLSCEASGHPRTISMVNGTA 470
Db 401 HKEVGLKRRKTKLIVEGRP--QIKMTKKTNTNKKMKTIVCHVEGFPKPAQVITGSG 458
Qy 471 S--EODODPQVRLSTLN--VLVTPLELLETGVECTASNDLGKNTSILFLELVNLTLLPDS 526
Db 459 SLINKTEETKYVNGKFSKRIIAPENVT-LTICIAENEL-----E 497
Qy 527 NTTGLSTAS-----PHTRANSTSTERKLPPEPSRGVIVAVIYVILVAVLAV 580
Db 498 RIVTSLNVAISAIPIEYDEPEDR--NDONSEK---VNDQAKLIVGIVGLLVALVAVY 552
Qy 581 FLY-KKGLPCRRSGKEITLPPSRKSE 607
Db 553 WLTVKSKTASKAVHDKDLGNIEENKLE 580

RESULT 7
JH0464
DM-GRASP precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 02-Sep-2000
C:Accession: JH0464
R:Burns, F.R.; von Kannen, S.; Guy, L.; Raper, J.A.; Kamholz, J.; Chang, S.
Neuron 7, 209-220, 1991
A:title: DM-GRASP, a novel immunoglobulin superfamily axonal surface protein that sup
A:reference number: JH0464; MUID:91337449
A:Accession: JH0464
A:Molecule type: mRNA
A:Residues: 1-587 <BUR>
A:Experimental source: brain
C:Comment: This is a cell surface glycoprotein.
C:Comment: This protein is localizes to axons in the dorsal funiculus and ventral mid
C:Keywords: glycoprotein
E:1-32/Domain: signal sequence #status predicted <SIG>
E:33-587/Product: DM-GRASP #status predicted <DMG>
F:67,198,270,311,365,461,484,503/Binding site: carbohydrate (Asn) (covalent) #status

Query Match 14.9%; Score 502.5; DB 2; Length 587;
Best Local Similarity 26.3%; Pred. No. 1.2e-24;
Matches 165; Conservative 118; Mismatches 254; Indels 91; Gaps 30;

Qy 6 LVCAFLAACCCCPRVAGVGEAEQAPAEVVEVGSTALLKCGLSQSGNLSHYDW-FS 64
Db 17 LLC--LLLAALCMPRALG-----YVNAVYGDITMPCRLVEPDG-LMEGKMKYE 64
Qy 65 VHKERTLI-FR--VRQGGGSEPEYEORLSLQDRGATLALQVTPQDERIFLCO--GK 119
Db 65 MPNGSPVFIAPFRSTKKNVQYDDVYDKRLSLSE--NYTSLIKNARIHDEKRFVCMVTE 123
Qy 120 RPRSOEYRIQLVYKAPPEPNI--QVNPILGIPVNSKEPEEVATCGVGRNGYPIPOVIWYKN 177
Db 124 DVVSEEPYV-VKVFQPSPELHQADFL----ETEKLMKLGECVVRDSYFPEGNTVWYKN 178
Qy 178 GRPLK--EENRNVIHQSSQVYESSGLYTLQSLKAQLYKDKDAQFCELYNRLPSGNHM 235
Db 179 GRVLQPEVEEVVNLNRKEN--RSTGLFTMTSLOYMPTKEDANNAKFTCIYTHGSPSQKT 237
Qy 236 KESREYTVAVFPTEKVMLEV--EPVGMLEKEDGRVETICLADGNPPH--FSISKQNPST 291
Db 238 IQSEPVFVDVHYPTREKTIIVLSSQSTIKRGDNVTLKCSGNGNPPQEFLEYI----- 290
Qy 292 REAEETTNQNGVLVLEPARKHSGRYECQGLDITMTLSLSEPOLLVNVYSDVRSVA 351
Db 291 -PGETGIRSSDYVMTDVRNRATGEGYKCSLIDKSM-----DATITVHYL-DLQLPFS 343
Qy 352 AP-EROGSSSLITLCEASSQDLFEQWLREFTGVLEKRPVLQDLKREAGGRCVNS 410
Db 344 GEVTKQIGALPVSCTISSSRNATVFWIKDNTR--MKTSP--SFSSLOYODAGNICETT 399
Qy 411 VPSIPLNRTOLVNVALFPGPMMAFKERKVVYKENVNLNLSCEASGHPRTISMVNGTA 470

Db 400 LOEPEGJMKKTKTLKIVEGRP--QIKMTKNTNKKSKITLVCHVEGFPRPAWQMTVTS 457
OY 471 S--EODDDPQRVLSTLN--VLVTPPELLTGVECTASNDLCKNTSLIFLELVNLTITTPPS 526
Db 458 SLINKTEETKYVKGKTSKITTIAPEENV--LTCTIAENL-----E 496
OY 527 NTTGGLSTAS-----PSTRANSTSTERKLPEPESRGVIAVAVICITLAVLGA 580
Db 497 RVTSLNVAISLIPYDEPEDR--NDNSEK---VNDQAKLIVIGVGLLVALVAGVY 551
OY 581 FLV--KKGKLPGRSGKQETLPPSRSE 607
Db 552 WLTVKSKTSKSHVDRDLGNIENKLE 579

RESULT 8

I39428
alcam - human
C:Species: Homo sapiens (man)
C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
C:Accession: I39428
R:Bowen, M.A.; Patel, D.D.; Li, X.; Modirell, B.; Malacko, A.R.; Wang, W.C.; Marguardt, H.
J. Exp. Med. 181, 2213-2220, 1995
A:Title: Cloning, mapping, and characterization of activated leukocyte-cell adhesion mol
A:Reference number: I39428; MUID:95279947
A:Accession: I39428
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-583 <RES>
A:Cross-references: GB:I38608; NID:9886257; PID:AA859499.1; PID:9886258

Query Match 14.6%; Score 492.5; DB 2: Length 583;
Best Local Similarity 24.8%; Pred. No. 5.2e-24;
Matches 145; Conservative 121; Mismatches 256; Indels 63; Gaps 23;

OY 41 GSTALLKCGLSQSGNLSHVDFSVHKEKRTLIFVRQGGOGS---EPGEYORLSLOD 96
Db 36 GDTIIIPCRLDVPRQ--NLMEGKMKYKPKDGSPIFARSSSTKSSVQYDDVPEKDRNLSE 94
OY 97 RGTALTOVTQDERIFLQ--GKRRSGEYRIQLRVYKAPBEPIQVNPGLGIPVNSKP 155
Db 95 -NLTLSISNARISDEKRFVCMVTEEDNVEAPTIYKVFQPSKPELVSKALFL--ETEDL 151
OY 156 EEVATGVGRNGYPIPOVIYKNGRPLKEEKNRVHIQSSQTVES--SGLYTLOSLTAKQLV 214
Db 152 KRLGDCISEDSIPDGNITTYRNGKVLHPLEGAVIIFKEMDPVOLTMTSTLEKTKR 211
OY 215 EDKDAQFYCELNRYLPSGNHMKESRETVVPFVPEKVLVLE--EPVGMLEKEDRAVEIRCL 273
Db 212 ADIQMFQTSVLYYSGSGKTIHSEQAVFDIYPTQVYIQVLPRNMAKEDNDITTKL 271
OY 274 ADGNPPH--FSISKQNPSTREAEETTNDNGVLVLEPARKEHSGRECOGLDITMTS 330
Db 272 GNGNPPEPEFLYLPQAPGIRSSNYYTLM-----VRNNAAGDYCSLIDKSMIA 323
OY 331 LTSEPQELLVNVSDRVSPAP--EROEGSSLTITCEASSODLEFQWIRETGVLERG 389
Db 324 STA-----ITVHTL-DLSLPSEGVTRQIDALPVSCTTASASNAIVVWKKDIR--LRSS 376
OY 390 PVLQHLDLKREAGGYRCVAVSPISGLNRTOLVNAVIFGPWMAFEKKEKRVVVKENMYLN 449
Db 377 P--SFSSLIHYQAGNVCCTALQEVGELKKRSLTLIVGKR--QIKMTKTRDPGSLSKT 432
OY 450 LSCFASGHRPRTISMNVTAS---EODDDP--RVLSTLVNLTPELLETGVECTASN 503
Db 433 IICHVEGFKPAIOWITITSGSVINQTESPYINGRYS--KIISPEENV--LTCTAEN 489
OY 504 DKGKNTSLIFLELVNLTITLTPDSNTTGLSTASPHRANSTSTERKLPEPESRGVIV 563
Db 490 QLERTVNSLNVSAISI--PE-----HDADETSIDENR-EKVNQDAKLIV 530

OY 564 AVIYCIILVAVLGAIVFLY--KKGKLPGRSGKQETLPPSRSE 607
Db 531 GIVVGLLALVAVGAVVYIMLMKSKTSKSHVKNKDLGNNEKLE 575

RESULT 9

150478
neurolin - goldfish (fragment)
C:Species: Carassius auratus (goldfish)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C:Accession: 150478
R:Jäessing, U.; Giordano, S.; Stecher, B.; Lottspeich, F.; Stuermer, C.A.
A:Title: Molecular characterization of fish neurolin: a growth-associated cell surface
protein DW-GRASP/SC-1/BEN.
A:Reference number: 150478; MUID:94299040
A:Accession: 150478
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-523 <LAE>
A:Cross-references: GB:I25056; NID:g407318; PID:g407319

Query Match 13.6%; Score 457; DB 2: Length 523;
Best Local Similarity 26.7%; Pred. No. 8.3e-22;
Matches 158; Conservative 108; Mismatches 226; Indels 100; Gaps 25;

OY 41 GSTALLK--GLSQSGNLSHVDFSVHKEKRTLIFVRQGGOGS---SEPEYORLSL 94
Db 2 GETIIVPCNDGTKRPDQ--LIPTKMKYVVDGSPDGLVKKQKDEATVATDGYKSHVSI 60
OY 95 QDRGATLALQVTQDERIFLQ--GKRRSGEYRIQLRVYKAPBEPIQVNPGLGIPVNSK 153
Db 61 A-ANSSLLIARGSLADQVFTQVNVSTNLEYSVEVKHKKPSAPVKK-----NNAK 112
OY 154 EPE-----EVATGVGRNGYPIPOVIYKNGRPLKEEKNRVHIQSSQTVES--SGLYTLOSI 207
Db 113 ELENKTLQLECEYVKNANPPADLWKKNNOTLVDDGKTIITSTIKDKITGLSTSSR 172
OY 208 LKAQIVKDKDAQFYCELNRYLPSGNH--KESRETVVPFVPEKVLVLEPVGMLK 263
Db 173 LQYARKKEDVESQFTC-----TAKHVMGPDQVSESESPFIHYPTREKVSIGVVSQSPIR 225
OY 264 EGDREITCLADGNPP--HESSISKQNPSTREAEETTNDNGVLVLEPARKEHSGREYEC 320
Db 226 EGEDVTLKQADGNPPSPSENFNI-----KKKKVTYDKDVTYTLGVTRADSGIYKC 277
OY 321 QGLDITMISLSEPOELLVNVSDRVSPAP--ROEGSSLTITCEASSODLEFQWIR 379
Db 278 SLINDVNES-----IQFVTVSFL-DVSLTPGKVLKAVGENLIVSLDKNASSSEAKYVWK 332
OY 360 EETGVLERGVLOLHLDKREAGGYRCVAVSPISGLNRTOLVNAVIFGPWMA--FKE 437
Db 333 D--NRKIDKLDP--DFSKLTYSDAGLVYCDV---SIEIKRSLSPFLVEGSIPTITSYTKH 385
OY 438 RKVWYKEMVNLSCFASGHRPRTISMNVTASQDQDDQVYVSTLVNLTPELLETGV 497
Db 386 RSSQKHKV--LTCEAGSGSPKPDVQMSVNGTNDVSYNGKATYKTLTVPSKNLT--V 439
OY 498 ECTASNDLGKNTSLIFLELVNLTITLTPDSNTTGLSTASPHRANSTSTERKLPEPES 557
Db 440 SCLVTNKGEDTKRI-----SVFSQKNKEDTE----- 466
OY 558 RGVYIAVAVICITLAVLGAIVFLY--KKGKLPGRSGKQETLPPSRSE 607
Db 467 QAKVIVGIVGLVLAALVGLVITWYIKTRQSGWMKGEKEAGTSSESKLE 518

RESULT 10
T20992
hypothetical protein F15G9.4a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T20992; T24733
R:Stulson, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
C:Accession: T20992
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5175 <WIL>
A:Cross-references: EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:FL5G9.4a
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
C:Accession: T24733
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5175 <WIL>
A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:FL5G9.4a
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: CESP:FL5G9.4a
A:Map position: X
A:Intons: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
A:Truncs: 253/3; 269/3; 275/1; 285/1; 288/3; 291/3; 294/1; 296/3; 299/3; 303/1;
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 5077/

Query Match 8.4%; Score 281; DB 2; Length 5175;
Best Local Similarity 22.1%; Pred. No. 3.6e-09;

Matches 122; Conservative 91; Mismatches 220; Indels 118; Gaps 27;

QY 27 EAQOPAPAELEVEVGSTAL-----LKCGLSOSOGNLSHVMFVHKERTLIFRVROG- 79
DB 1357 DVQEPPIILPSTQTNNAVGDVRELKCYEASPP--ASTWTF-----RRGI 1401
QY 80 -QGQSEGEYEQRLSLDRGATLALQVTPQDERIFLCQGRPRSQ--EYRIQLRYKAPE 137
DB 1402 AIGDTGTG-----YVESDGTLVIOSASVEDATITTCASNPAGKAEANLQVTVIASPD 1455
QY 138 --EPNIQVPLGIPVNSKEPEVATCGRNGYPIPOVIWYKNGRPL-----KEKN 186
DB 1456 IKDDVVTQD---SIKSHPSFLYCPVFSN--PLQIIMYLNDKRLIDDKTSWKTSDDKR 1510
QY 187 RVHIQSSQTVESGLYTLQSLKAOLVKEDKDAQFYCELYRLPSGNHMKESR-EVTVPV 245
DB 1511 KLHVFRAK-ITDSGVYKCA-----RNAAGEGSKSFQVEVTVPL 1548
QY 246 FYFTEKMLEVEPVGMLEKGDVREIKCLADGNPPHFSI-----SKONPSTRAEETND 301
DB 1549 NLDESKYKKRV---FAKEGEVTLGCPVSGFPVQIIMVVDGTVVEPGKKYKATLSND 1604
QY 302 NGVLVLEPARKHSGRYECOG-----LDLDTMISLSEPOELLVNVYSDVRVSPAPER 355
DB 1605 GLTLHFDSVSVKQEGNHCVAOSKGNLIDIVELSYLAVP---IYGEDDNLVFP----- 1655
QY 356 QEGSSLTITC--EAESSQDLEFQWL-----REETGOVLERGPLYQLHDLKRBAGGGR 406
DB 1656 -LGKDISLSCDLQTESDDKTFVWSINGESDRPDNVQIPSDGHRLYITDAKPNNGKYM 1714
QY 407 CVASVPSIPIGLNTQLVNV---AIFGPPMAFKERKVMVKNVNLNLSGASGHPRPITS 463
DB 1715 CRYI-NSAGAERTLTLDVLEPPVFER--VEANOKLIGNNPII-LQCCVTSNPPRYAI 1770
QY 464 MNVNGTSEEDODPQRYLSTFLNVLPPELLETG---VECTASNDLGKNTSILFLVNL 519
DB 1771 WKIDGNVDKSWLFDSELSLRT-----EKL-TKSAOISCTAENKAGTASRDFFIONIA 1825
QY 520 TTITPDSNTTT 530
DB 1826 PTFKNEGDEET 1836

RESULT 11

T43290
hemicefelin precursor - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text-change 18-Feb-2000
C:Accession: T43290; T20993; T24734
R:Voegel, B.E.; Hedgecock, E.M.
submitted to the EMBL Data Library, June 1998
A:Description: Hemicefelin is required for hemidesmosome mediated cell adhesion and ge
A:Reference number: Z22396
A:Accession: T43290
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5198 <VOG>
A:Cross-references: EMBL:AF074901; PIDN:AAC26792.1
R:Stulson, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19355
C:Accession: T20993
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <WIL>
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CESP:FL5G9.4b
A:Experimental source: clone F15G9
R:Kershaw, J.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19929
C:Accession: T24734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5198 <WIL>
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CESP:FL5G9.4b
A:Experimental source: clone T09B9
C:Genetics:
A:Gene: him-4; FL5G9.4b
A:Map position: X
A:Intons: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;
A:Truncs: 253/3; 269/3; 275/1; 285/1; 288/3; 291/3; 294/1; 296/3; 299/3; 303/1;
1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 51

Query Match 8.4%; Score 281; DB 2; Length 5198;
Best Local Similarity 22.1%; Pred. No. 3.6e-09;

Matches 122; Conservative 91; Mismatches 220; Indels 118; Gaps 27;

QY 27 EAQOPAPAELEVEVGSTAL-----LKCGLSOSOGNLSHVMFVHKERTLIFRVROG- 79
DB 1357 DVQEPPIILPSTQTNNAVGDVRELKCYEASPP--ASTWTF-----RRGI 1401
QY 80 -QGQSEGEYEQRLSLDRGATLALQVTPQDERIFLCQGRPRSQ--EYRIQLRYKAPE 137
DB 1402 AIGDTGTG-----YVESDGTLVIOSASVEDATITTCASNPAGKAEANLQVTVIASPD 1455
QY 138 --EPNIQVPLGIPVNSKEPEVATCGRNGYPIPOVIWYKNGRPL-----KEKN 186
DB 1456 IKDDVVTQD---SIKSHPSFLYCPVFSN--PLQIIMYLNDKRLIDDKTSWKTSDDKR 1510
QY 187 RVHIQSSQTVESGLYTLQSLKAOLVKEDKDAQFYCELYRLPSGNHMKESR-EVTVPV 245
DB 1511 KLHVFRAK-ITDSGVYKCA-----RNAAGEGSKSFQVEVTVPL 1548
QY 246 FYFTEKMLEVEPVGMLEKGDVREIKCLADGNPPHFSI-----SKONPSTRAEETND 301
DB 1549 NLDESKYKKRV---FAKEGEVTLGCPVSGFPVQIIMVVDGTVVEPGKKYKATLSND 1604
QY 302 NGVLVLEPARKHSGRYECOG-----LDLDTMISLSEPOELLVNVYSDVRVSPAPER 355
DB 1605 GLTLHFDSVSVKQEGNHCVAOSKGNLIDIVELSYLAVP---IYGEDDNLVFP----- 1655
QY 356 QEGSSLTITC--EAESSQDLEFQWL-----REETGOVLERGPLYQLHDLKRBAGGGR 406
DB 1656 -LGKDISLSCDLQTESDDKTFVWSINGESDRPDNVQIPSDGHRLYITDAKPNNGKYM 1714
QY 407 CVASVPSIPIGLNTQLVNV---AIFGPPMAFKERKVMVKNVNLNLSGASGHPRPITS 463

Db 1715 CRVT-NSAKAERTLTLDVLEPEVPEP--VEEANOGLIGNPIL-LQQVQVGNKPEVTI 1770
OY 464 WNNNGTASEODODPOKRVSLTANLVLPPELTG----VACTASNDGKNTSILFELVNL 519
Db 1771 WKIDGNDVDKSWLFPDESLILRI-----EKL-TGKSAQISCTAENKAGTASRDFQIONIA 1825
OY 520 TTLTPDSNTT 530
Db 1826 PTFKNEGDOET 1836

RESULT 12

A49448
Irregular chiasm C-roughest precursor - fruit fly (Drosophila melanogaster)
N:Alternate names: irrec-roughest protein
C:Species: Drosophila melanogaster
C:Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 07-May-1999
C:Accession: A49448; S34129
R:Ramos, R.G., P.: Iqbal, G.L.; Lichte, B.; Baumann, U.; Maier, D.; Schneider, T.; Brandst
Genes Dev. 7, 2533-2547, 1993
A:Title: The irregular chiasm C-roughest locus of Drosophila, which affects axonal proje
A:Reference number: A49448; MUID:94102535
A:Accession: A49448
A:Status: preliminary
A:Molecule type: DNA; mRNA
A:Residues: 1-764 <RAM>
A:Cross-references: GB:LI1040; NID:9304790; PID:9304791; EMBL:Z21641; NID:9312985; PID:9
A:Gene: flyBase:rst
A:Cross-references: flyBase:FBgn0003285
C:Keywords: transmembrane protein

Query Match 8.0%; Score 270; DB 2; Length 764;
Best Local Similarity 20.7%; Pred. No. 1.3e-09;
Matches 137; Conservative 89; Mismatches 277; Indels 158; Gaps 25;
OY 40 VGSTALLKCGLSQSGNLSHV--DMFSVHKEKRTLIFRVNGGQSGSPGEYBGRSLSDNG 98
Db 41 VGARVTLPRVINKOGTLMTKDDFGLTSDLSGFERAYAWGSDDEGY----- 90
OY 99 ATLALQVTPDERIFLCQCK-----RPRSOEYRIQLRYKAPPEPNIDVNLGIPVNSK 153
Db 91 -SLDIYPMALDDARQOCQVSPGEGPAIRSTFAGLTIVLPPEAKPTIGQVITYATADR 149
OY 154 EPEEVAATCVGRNGYPIPOYIWKV--GR-----PLKEKNKRVHIOSSQTVSSGL 201
Db 150 KVE--IECVSVGKPAAEITWIDGLGNVLTDNIEYTVIPLPDQR----- 192
OY 202 YTLQSLTKAQVKEKDAQFYCELNRLPSGNHMKESREVTYPVYPTIEKYVLEVEPYGM 261
Db 193 FTKASVLRILPKKEHNHNFSCQAQ--NTADRTYRSARIRKRYVYAP--VKVNVKMS 246
OY 262 LKEG-----DRVEIRCLADGNPP--HFSISKNPSTREAE 295
Db 247 LRGAGAGSVGAGGSGVHMSSTGRIVEHSQVLECRADANPDSVYKRWFINDEPITIGGK 306
OY 296 EETTDNGVLYLEPARKHSGREKCGLDLDTMISLSEPOELLVNVSDVVSAPAEPR 355
Db 307 TE-----WVINRVTKRFHDAIVKC--EVQNSVGKSEDSFTLIDISYASFQRORSMDA 357
OY 356 QEGSSSLTLCESASODLEFQWLREFTGOVLERGPVLQHLHLKRAAGGYRCVAVSPISF 415
Db 358 DVGGSVSLTCEVDSPQPIVMIQHPDSRVVGTSTNLT--SVSNFTAGYCYKAVPGYA 416
OY 416 GLNRQLVNVAFGPPEMAFKERKYVVENVNLSCASGHPRTI--ISMVNG--TASE 472
Db 417 EISADAVYVTK--GSPAIG--SQRTOYGLVGDARIECFRASSVPARRHVSMTNGGEISSE 473
OY 473 QDOD-----PQVYSLTANLVLPPELTGECTASNDLGNKTSILFELVNLTLTP 524
Db 474 SGHDYSILVDAVPGVSKTLIRDSQAYHYGKYNCTVVDYNDVAEIQLAKKKSVSL-- 531

OY 525 DSNITTGSLTASPHRTANSTIERKLEPEPSGVIVANIVICILVAVLAVLYFLYK 584
Db 532 -----LMTIVG-----ISVAVLVLTLTL-VVVY----- 555
OY 585 KGKLPGRSGROEITLPPSRKSELVEYKSDKLEPMGLQSSGDRAPDQGEYIDL 644
Db 556 ---IKCKRRK-----LPP-----ADVISEHQITKNGVSCGLEFDGRISN;SDL 597
OY 645 R 645
Db 598 K 598

RESULT 13

JC2457
vascular cell adhesion protein - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 15-Feb-1995 #sequence_revision 05-Apr-1995 #text_change 08-Oct-1999
C:Accession: JC2457
R:Tseng, Y.T.M.; Haskard, D.O.; Robinson, M.K.
Biochem. Biophys. Res. Commun. 201, 805-812, 1994
A:Title: Cloning and expression kinetics of porcine vascular cell adhesion molecule.
A:Reference number: JC2457; MUID:94271236
A:Accession: JC2457
A:Molecule type: mRNA
A:Residues: 1-538 <TSA>
A:Cross-references: EMBL:U08351; NID:9474382; PIDN:AAA21542.1; PID:9474383
C:Keywords: glycoprotein; transmembrane protein
F:497-517/Domain: transmembrane #status predicted <TM>
F:75,157,271,330,360/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 7.9%; Score 264.5; DB 2; Length 538;
Best Local Similarity 21.6%; Pred. No. 1.8e-09;
Matches 126; Conservative 103; Mismatches 234; Indels 119; Gaps 27;
OY 31 PAPELVEYEVGSTALLKCGLSQSGNLSHVDWFSVHKEKRTLIFRVNGGQSGSPGEYEQ 90
Db 31 PEDKMI-AQIDSDASLTCSADPCSSLS-FSW-----RTQIDSPUNG----- 70
OY 91 RLSLQDRCATLALQVTPDERIFLC-----QGRPRSOEYRIQLRYKAPPEPNIQV 143
Db 71 KVTNGTSTILVMPVSEFNHSLTCTVSGCNLKER-----GIQVEITYSPRPDPTIW 124
OY 144 NPL---GIPVNSKEPEEYATCVGRNGYPIP--QVIWYKNGRPKLEEK--NRVHIOSSQTV 196
Db 125 SSLPEVCGKPVTVR-----CLVPDQVYVPEKLEIELLKDNHMSVQSQFLELIDIKSETEK 177
OY 197 ESSGLYTLQSLTKAQVKEKDAQFYCELNRLPSGNHMKESREVTYPVYPTIEKYVLEV 256
Db 178 SLEFTFT-----PTEEDIGKAIVCOATLIDGQSVVTTPEKMOVYISPPDPV-ISV 228
OY 257 EPGVGLKRGDEVEIRCLADGNPPHFSISKNPSTREAEETTDNGVLYLEPARKHSG 316
Db 229 NPSTSLQDGDMMATCTSEGLIPAFQISMSK--LDNGDQDLISGNAFTLLIMRMEDSG 285
OY 317 RYECQGLDLMTMSILSEPOELLVNVY---SDVVSAPAEERQSGSLTLCESASODL 373
Db 286 IYVEGVN---PVGITNKREVELVQVAPRDTTISVNSS--TLEGSSVNNMTCSSDGPAP 341
OY 374 EFQW---LREFTGOVLERGPVLQHLHLKRAAGGYRCVAVSPISPGINRQLVNVAFICP 430
Db 342 KILMSKRLRDGNLPLISENTTLTLTSTRMDSGIVYCEGI--NOAGINRKE-VELIIQAA 398
OY 431 P-----WMAFKERKYVVENVNLSCASGHPRTITIMVNVNGTASRODODPOKRVSLTANV 486
Db 399 PKDQLTAFPSBSKBEQDYIYSTC---GNVPTTL-----ILKKAETGDIYVLSSTG 450
OY 487 LVT---PELLETV-ECTASNDLGNKTSILFELVNLTLTPPSNTTGTGSLTASPHTR 542
Db 451 AYTIHRRARLADAGVYECESKMEIG-----LQLSRITLDVAGRSN----- 490

OY 543 ANSTERKLEPEPESRCVIAVAVICLVLAVLAVLYFLYK 584
Db 491 KDYFSS-----LVLVLCASSLIIPALGVITTFARK 521

RESULT 14

A39712
kinase-like protein klg precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 08-Nov-1991 #sequence_revision 08-Nov-1991 #text_change 24-Sep-1999
C/Accession: A39712
R:Chou, Y.H.; Hayman, M.J.
Proc. Natl. Acad. Sci. U.S.A. 88, 4897-4901, 1991
A:Title: Characterization of a member of the immunoglobulin gene superfamily that possib
A:Reference number: A39712; MUID:91271300
A/Accession: A39712
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1051 <CHO>
A:Cross-references: GB:M63437; NID:q212235; PIDN:AAA48933.1; PID:q212236
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C:Keywords: ATP
F:775-1046/Domain: protein kinase homology <KIN>
F:783-791/Region: protein kinase ATP-binding motif

Query Match 7.7% Score 259.5; DB 2: Length 1051;
Best Local Similarity 20.1%; Pred. No. 9.4e-09;
Matches 149; Conservative 89; Mismatches 263; Indels 241; Gaps 31;

OY 29 EOPAPLEVEVEGSTALLKCGLSQSGNLSHVD-----WF-----SVHK 67
Db 119 KQPA-SAAEIQPSSTVVLRC-----HIDGHPRTQMFRDGAFLPDGRGTYSVS 167
OY 68 EKRLIRRVAGOGQSGPEY-----EQLSLQDKRATIALTQ-----V 106
Db 168 KERLLTLR---GAGPDNGILYGCARPRAVGVSCSDNFTLITIDSEFQAVVVPEDLV 224
OY 107 TPODERIFLCO--GKRPSOEY-----RIQUR---VYK--- 134
Db 225 TKNEAMFDQFAVPPPTQEMLEFDSPTNRKSTTFANGSLITLVKARKSGVYKCI 284
OY 135 -----APEPNIOVNLGITVNSKEBEVATCVGRNGYPIPOV 172
Db 285 HGCKKALVLKATLRLAEIEMAFSPKVLTLANOGHRV-----SCACPGVPTPOV 335
OY 173 IWKY-----GRPLKEKNRVHIOSSQTVESGLYTLQSLIKALQVKDKDAQFCYL 225
Db 336 WMEKNQERVPAGRVYQEAQLV--TSTIEADAGITTCMAANKA----- 378
OY 226 NYRLPSGNHMKESREVTVPVYPTKVMLEVEPYGMLKEGDREIRCLADGNPPHFSIS 285
Db 379 -----GEKKQELSTIV---ATVPKVMVEPKDQLEESKPGYHLCKSLPTVTWY 427
OY 286 KQNSTEAEETINDNGVLLVLEPAKKEHSGRYECQGLDITMSLSLSEFQELLVY--- 342
Db 428 RNVGISIEDSRHFEISENGTLRINNVYVDGTMYKC-----VSTPAGSIEGARV 477
OY 343 --VSDVRSPAPAPRQ---EGSSLTIFCEASSODLEFQMLREB---TCQVLERGVLQ 393
Db 478 HVLKELKFTPPROPLOQMEFKKVTYVCSATGRKPTIOMTKRIDGSSLPESHVSRAGILS 537
OY 394 LHLDKREAGGCGYRCVAVSPISIPGLNTQRLQVNLVAFGPPMAFK--ERRVVKENMVLNLS 451
Db 538 FHAKSRSDSCNYCTIASNSPGGEIRATVQLVAVY---VTFLKEPPTVYVCGHTAMFQ 593
OY 452 CEASGHPPTISWVNGTASQDODPORVLSTLNLVLTPELLETGVCTASNDLGKFTSI 511
Db 594 COAEGDVPPIQMK-----GKDKILDPKSLPLRIQIMPGSLV---IYDVTTEDSGKYTCI 646
OY 512 -----LPELVNLTLTPDSDNTTGLSTASTASPTPRANSTERRKLEPESRGV 560
Db 647 AGNSCNITKHRFAFLVYVDKPAAEDEG-----PSSHTPYKMIOT-----TGL 688

OY 561 VIVAVICLVLAVLAVLYFLYKKGKLPKRSGKQETLPPSRKSEL----- 608
Db 689 SYGAAYVITIT--VLG--LMFYCKR-----RRKAKRLKKHDEGEPEHECLNGTLLQNG 739

OY 609 --VVEKSDRIPEENGLLQSSS 628
Db 740 QTTAEIQEVALTNLGGSSGAS 761

RESULT 15

LUXNL
neural cell adhesion molecule long domain form precursor - African clawed frog
N:Alternate names: NCAM-180
M:Contains: neural cell adhesion molecule, short domain form (NCAM-140)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 22-Jun-1999
C/Accession: S09600
R:Krieg, P.A.; Sakaguchi, D.S.; Kiltner, C.R.
Molecular Acids Res. 17, 10321-10335, 1989
A:Title: Primary structure and developmental expression of a large cytoplasmic domain
A:Reference number: S09600; MUID:90098871
A/Accession: S09600
A:Molecule type: mRNA
A:Residues: 1-1088 <KRI>
A:Cross-references: EMBL:M25696; NID:q214609; PIDN:AAA49909.1; PID:g214610
A:Note: the authors translated the codon AAA for residue 970 as Leu
C:Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM
C:Comment: Several forms of NCAM are produced by alternative splicing.
C:Genetics:

A:Gene: NCAM
C:Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; I
C:Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding;
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1088/Product: neural cell adhesion molecule, long domain form #status predicted
F:20-803/1050-1088/Product: neural cell adhesion molecule, short domain form #status
F:20-705/Domain: extracellular #status predicted <EXT>
F:34-95/Domain: immunoglobulin homology <IMM1>
F:129-188/Domain: immunoglobulin homology <IMM2>
F:149-153/Region: heparin binding #status predicted
F:158-162/Region: heparin binding #status predicted
F:225-284/Domain: immunoglobulin homology <IMM3>
F:317-381/Domain: immunoglobulin homology <IMM4>
F:413-475/Domain: immunoglobulin homology <IMM5>
F:512-589/Domain: fibronectin type III repeat homology <FN3A>
F:618-679/Domain: fibronectin type III repeat homology <FN3B>
F:706-723/Domain: transmembrane #status predicted <TM>
F:724-1088/Domain: intracellular #status predicted <INT>
F:41-93,136-186,237-282,323-379,420-473/disulfide bonds: #status predicted
F:219,310,341,417,443,472/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 7.6% Score 256; DB 1: Length 1088;
Best Local Similarity 18.1%; Pred. No. 1.7e-08;
Matches 155; Conservative 100; Mismatches 266; Indels 336; Gaps 35;

OY 33 PELVEVEGSTALLKCGLSQSGNLSHVDSEVHKERTLFRVAGOGQSGPEYEDRL 92
Db 26 PQGGEISLSESKFFLC---QVSGEATDISWSPTEKILVT-----QOOI 66
OY 93 SL---QDRATLALQVTPQDERIFLC-----QGR-----RPRSOEY 126
Db 67 SYVRSDDYVSTLTYINASSODAGITCVASNEAGESEGTYNLKYOLTKRNAFTPOEF 126
OY 127 R-----TOLRVYKAPR----- 138
Db 127 KEGEDAVIICDVSSSITITRRHKKQVIEKKDVRFLVANNVLIQINGIKKTDCGTIRC 186
OY 139 -----PNIQVNLGLIPVNSKEPEVATCVGRNGPIPOVITYK 176
Db 187 EGRILARGEINVKDIQVIVNVPPTIARQLRVNANNAEVSVLSCDADGEPDEPISWLK 246
OY 177 NGRPLKEKNRHHIOSSQTVESGLYTLQSLIKALQVKDKDAQVYCELNTRLPSGNHMK 236

```

Db 247 KGEPLDGEEEKISFNED-----QSEMTIHHEKDEADEYSCIAN-----NQAG 289
QY 237 ESREYTVFVPTREKVMLEVEPGMLKEGDRVEIRCLADNPPPHFISIKONSTREAE 296
Db 290 EAEATILIKYAKKITYVENKTAV--ELDEITLTCEASGDDIP--SITWRTAVRNTISSE 345
QY 297 ETTNDNGVLVLEPARKE-----HSGRYEC-----OGLDIDTMTISLSEFOELVN 341
Db 346 ATTLDDGHIIVKVEHIRMALTLKDQYTDAGEYFCIASNPIGVMQAMT-----FEVQ 397
QY 342 YSDVVRVSPAAPERQESSFTLTCEASSODLEFQWLREFTGOVLE-----RGPV 391
Db 398 YAPKIR-GPVVVVYTWEGNPNVITCEVFAHPRAAVTWFRD--GOLLPSNFSNITIKISGP 454
QY 392 ---LQLDLREAGGGRV-----ASVPSIFGLNRTQLVN--VAIFG 429
Db 455 SSSLEVPDSENDENGNCTAINTIGHSEFFILVQADTPSSPAIRKVEPYSSTVMIVFD 514
QY 430 PP-----WMAFERKRVWVK-----ENM--VLNLSCEAS-----455
Db 515 EPDSTGVPLIKYKAEMRVIGHEKMHKYYDAKEVNAESIITVMGLKETSYWKLAMN 574
QY 456 ----GHPRTISMNVN-----GTASE-----QODDPQRV-----LST 483
Db 575 GKGLGDSTPSQEFITQVREPSAPKLVGHLSDEGNSIKVDILKODDGGSPIRHYLVNRA 634
QY 484 LNVLV--TPEL-----LETGYE-----CTASNDLGKNTSTL--FLELVNLTTLT 523
Db 635 LNALEKPEKRVPSNSHHVWLKALEWVDYEVIVAEENOQKSPALLSFRTAKPTATT 694
QY 524 PDSNTTGLSTASPHTRANSTSTERKLPESRGVIVAVICILVLAVALGYLYFLY 583
Db 695 ATASAGTGLGTG-----AIVGILIVIFVLLVVDVYTCFFLN 731
QY 584 KKGKLC-----RRSGKOETLLPPSRKSELVEVKSDDLPEMGLQ 626
Db 732 KCGILMCIAVFCGAKGPGAKGKDIEGKAAPSKDESK--EPIVEVTE-----778
QY 627 SSGDKRAPGDQGEKYID 643
Db 779 ---EERTPNHDSNQIE 792

```

Search completed: March 23, 2001, 12:31:19
 Job time: 42 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 23, 2001, 12:30:42 ; Search time 13.25 Seconds

(without alignments)

1574.489 Million cell updates/sec

Title: US-09-653-961-2

Perfect score: 3363

Sequence: 1 MGJPLVCAFLAACCPCPR.....SSGDKRAPDQGEKTYDLRH 646

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3327	98.9	646	1	MU18_HUMAN
2	647.5	19.3	628	1	LU_HUMAN
3	509.5	15.2	588	1	C166_CHICK
4	492.5	14.6	583	1	C166_HUMAN
5	488.5	14.5	583	1	C166_MOUSE
6	465.5	13.8	555	1	C166_CARAU
7	438	13.0	564	1	C166_BRARE
8	270	8.0	764	1	ICCP_DROME
9	259.5	7.7	1051	1	PTK7_CHICK
10	256	7.6	1088	1	NCAL_XENLA
11	246.5	7.3	725	1	NCAL_MOUSE
12	246.5	7.3	739	1	NCAL_RAT
13	244.5	7.3	858	1	NCAL_RAT
14	244.5	7.3	3707	1	PGM_MOUSE
15	242.5	7.2	853	1	NCAL_MOUSE
16	237	7.0	1092	1	NCAL_BOVIN
17	235	7.0	4393	1	PGM_HUMAN
18	235	7.0	1091	1	NCAL_CHICK
19	234	7.0	761	1	NCAL_HUMAN
20	234	7.0	848	1	NCAL_HUMAN
21	233.5	6.9	1115	1	NCAL_MOUSE
22	232.5	6.9	1260	1	CAML_MOUSE
23	230	6.8	739	1	VCAI_HUMAN
24	230	6.8	837	1	NCM2_MOUSE
25	228	6.8	811	1	FS22_DROME
26	228	6.8	873	1	FS21_DROME
27	226.5	6.7	1259	1	CAML_RAT
28	226	6.7	738	1	PECL_HUMAN
29	225.5	6.7	1257	1	CAML_HUMAN
30	225	6.7	702	1	CCEM_HUMAN
31	224.5	6.7	1070	1	PTK7_HUMAN
32	218.5	6.5	404	1	RAGE_HUMAN
33	215	6.4	1284	1	NRCA_CHICK

RESULT	1	ALIGNMENTS
ID	MU18_HUMAN	STANDARD: PRT; 646 AA.
AC	P43121;	
DT	01-NOV-1995 (Rel. 32, Created)	
DT	01-NOV-1995 (Rel. 32, Last sequence update)	
DT	01-NOV-1997 (Rel. 35, Last annotation update)	
DE	CELL SURFACE GLYCOPROTEIN MUC18 PRECURSOR (MELANOMA-ASSOCIATED ANTIGEN MUC18) (MELANOMA-ASSOCIATED ANTIGEN A32) (S-ENDO 1 ENDOTHELIAL-ASSOCIATED ANTIGEN) (CD146 ANTIGEN) (MELANOMA ADHESION MOLECULE).	
DE	MCAM OR MUC18.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=MELANOMA;	
RX	MEDLINE=90099368; PubMed=2602381;	
RA	Lehmann J.M., Riettmueller G., Johnson J.P.;	
RT	"MUC18, a marker of tumor progression in human melanoma, shows sequence similarity to the neural cell adhesion molecules of the immunoglobulin superfamily.";	
RT	Proc. Natl. Acad. Sci. U.S.A. 86:9891-9895(1989).	
RL	[2]	
RN	REVISIONS, SEQUENCE FROM N.A.	
RP	TISSUE=MELANOMA;	
RC	MEDLINE=93391384; PubMed=8378324;	
RX	Sers C., Kirsch K., Rothbacher U., Riettmueller G., Johnson J.P.;	
RA	"Genomic organization of the melanoma-associated glycoprotein MUC18: implications for the evolution of the immunoglobulin domains.";	
RT	Proc. Natl. Acad. Sci. U.S.A. 90:8514-8518(1993).	
RL	[3]	
RN	SEQUENCE OF 24-44; 98-112; 135-153; 240-260; 379-389 AND 460-478.	
RP	MEDLINE=94215196; PubMed=8162602;	
RA	Shih I.-M., Eider D.E., Speicher D., Johnson J.P., Herlyn M.;	
RT	"Isolation and functional characterization of the A32 melanoma-associated antigen.";	
RT	Cancer Res. 54:2514-2520(1994).	
RL	[4]	
RN	SEQUENCE OF 27-40; 98-112 AND 236-260.	
RP	MEDLINE=96136502; PubMed=8573133;	
RA	Barlin N., Frances V., Lesaulle G., Horschowski N., George F., Sampol J.;	
RT	"Identification of the S-Endo 1 endothelial-associated antigen.";	
RT	Biochem. Biophys. Res. Commun. 218:210-216(1996).	
RL	[5]	
RN	FUNCTION.	
RP	MEDLINE=94122526; PubMed=8292890;	
RA	Johnson J.P., Rothbacher U., Sers C.;	
RT	"The progression associated antigen MUC18: a unique member of the immunoglobulin supergene family.";	
RT	Melanoma Res. 3:337-340(1993).	
RL	FUNCTION: COULD BE AN ADHESION MOLECULE ACTIVE IN NEURAL CREST CELLS DURING EMBRYONIC DEVELOPMENT. ITS EXPRESSION MAY ALLOW MELANOMA CELLS TO INTERACT WITH CELLULAR ELEMENTS OF THE VASCULAR SYSTEM THEREBY ENHANCING HEMATOGENEOUS TUMOR SPREAD.	

P20273 homo sapien
P35969 mus musculu
P35329 mus musculu
P29533 mus musculu
P13688 homo sapien
Q63495 rattus norv
P35968 homo sapien
Q28173 bos tauru
Q03696 gallus gall
P70211 mus musculu
P97798 mus musculu

```

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN
CC -1- TISSUE SPECIFICITY: MAY APPEAR AT THE SURFACE OF NEURAL CREST
CC CELLS DURING THEIR EMBRYONIC MIGRATION. APPEARS TO BE LIMITED TO
CC VASCULAR SMOOTH MUSCLE IN NORMAL ADULT TISSUES. ASSOCIATED WITH
CC TUMOR PROGRESSION AND THE DEVELOPMENT OF METASTASIS IN HUMAN
CC MALIGNANT MELANOMA. EXPRESSED MOST STRONGLY ON METASTATIC LESIONS
CC AND ADVANCED PRIMARY TUMORS AND IS ONLY RARELY DETECTED IN BENIGN
CC MELANOCYTIC NEVI AND THIN PRIMARY MELANOMAS WITH A LOW PROBABILITY
CC OF METASTASIS.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.
CC -1- DATABASE: NAME-PROW: NOTE-CD guide CD146 entry:
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd146.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL: M29277: AAA20824.1: -
DR EMBL: M28882: AAA20932.1: -
DR EMBL: X68264: CAA48332.1: JOINED.
DR EMBL: X68265: CAA48332.1: JOINED.
DR EMBL: X68266: CAA48332.1: JOINED.
DR EMBL: X68267: CAA48332.1: JOINED.
DR EMBL: X68268: CAA48332.1: JOINED.
DR EMBL: X68270: CAA48332.1: JOINED.
DR EMBL: X68271: CAA48332.1: JOINED.
DR EMBL: X68272: CAA48332.1: JOINED.
DR INTERPRO: IPR003006: -
DR PFAM: PF00047: 19: 5.
DR Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;
KW Signal.
FT CHAIN 1 23 CELL SURFACE GLYCOPROTEIN MUC18.
FT DOMAIN 24 559 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 560 583 POTENTIAL.
FT DOMAIN 584 646 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 39 122 IC-LIKE V-TYPE DOMAIN.
FT DOMAIN 154 229 IC-LIKE V-TYPE DOMAIN.
FT DOMAIN 265 327 IC-LIKE C2-TYPE DOMAIN.
FT DOMAIN 358 414 IC-LIKE C2-TYPE DOMAIN.
FT DOMAIN 445 506 IC-LIKE C2-TYPE DOMAIN.
FT DISULFID 48 116 PROBABLE.
FT DISULFID 161 223 PROBABLE.
FT DISULFID 272 320 PROBABLE.
FT DISULFID 365 407 PROBABLE.
FT DISULFID 452 499 PROBABLE.
FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 467 467 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 508 508 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 518 518 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 527 527 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 544 544 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 646 AA: 71793 MW: F064A5DAE0BAEC6 CRC04:

```

```

Query Match 98.9%: Score 3327; DB 1; Length 646;
Best Local Similarity 98.9%: Pred. No. 9,3e-213;
Matches 639; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 121 PROSEYRIOLRVYKAPEEPNIQVNPGLGIPVNSKEPEVATVGRNGYPIPOVITWYKNGRP 180
DB 121 PROSEYRIOLRVYKAPEEPNIQVNPGLGIPVNSKEPEVATVGRNGYPIPOVITWYKNGRP 180
QY 181 LKEERNVHIQSSQTVSSGLYTLQSLKQOLVKEKDQAFYELNRYLPSCGHMKESRE 240
DB 181 LKEERNVHIQSSQTVSSGLYTLQSLKQOLVKEKDQAFYELNRYLPSCGHMKESRE 240
QY 241 VTPVVFPTERVWVLEVPVCGMKRGDRVEIRCLADGNPPHFSISKONPSTRAAEETTN 300
DB 241 VTPVVFPTERVWVLEVPVCGMKRGDRVEIRCLADGNPPHFSISKONPSTRAAEETTN 300
QY 301 DNGVLVLEPARKHESGREGCGLDLDTMISLSEPOELLVNVYSDVSPAPAREQEGSS 360
DB 301 DNGVLVLEPARKHESGREGCGLDLDTMISLSEPOELLVNVYSDVSPAPAREQEGSS 360
QY 361 LTLTCEASSODLEFQWLREFTGOVLERGPVYLQHLDKRKARGGRCVAVSPSTPLGNT 420
DB 361 LTLTCEASSODLEFQWLREFTGOVLERGPVYLQHLDKRKARGGRCVAVSPSTPLGNT 420
QY 421 QLVNVAITFGPQWMAFKERKVVYKENVNLNLSCEASGHPRTISMNVNGTASEDDQDFQV 480
DB 421 QLVNVAITFGPQWMAFKERKVVYKENVNLNLSCEASGHPRTISMNVNGTASEDDQDFQV 480
QY 481 LSTLVNVLTPPELLFTGVECTASNDLGKNTSLFLELVNLTLPDSTMTTGLSTVASPH 540
DB 481 LSTLVNVLTPPELLFTGVECTASNDLGKNTSLFLELVNLTLPDSTMTTGLSTVASPH 540
QY 541 TRANSTERKLPPEESGVYIVAVICILVAVLGVAVLYLYKKGKLPGRSKQKQITL 600
DB 541 TRANSTERKLPPEESGVYIVAVICILVAVLGVAVLYLYKKGKLPGRSKQKQITL 600
QY 601 PPSRSELVVEKSPDLPEEMGLLOGSSGDKRAPDQGEKTIIDLRH 646
DB 601 PPSRSELVVEKSPDLPEEMGLLOGSSGDKRAPDQGEKTIIDLRH 646
QY 601 PPSRSELVVEKSPDLPEEMGLLOGSSGDKRAPDQGEKTIIDLRH 646
DB 601 PPSRSELVVEKSPDLPEEMGLLOGSSGDKRAPDQGEKTIIDLRH 646

RESULT 2
ID LU HUMAN STANDARD: PRT: 628 AA.
AC P50895:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE LUTHERAN BLOOD GROUP GLYCOPROTEIN PRECURSOR (B-CAM CELL SURFACE
DE GLYCOPROTEIN) (AUBERGER B ANTIGEN) (F8/G253 ANTIGEN).
GN LU OR BCAM OR MSK19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
SEQUENCE FROM N.A., AND SEQUENCE OF 32-67 AND 182-203.
RC TISSUE-PLACENTA:
MEDLINE-95596337: Pubmed-7777537:
PARSONS S.F., MALLINSON G., HOLMES C.H., HOULIHAN J.M., SIMPSON K.L.,
RA Mawdy W.J., Spurr N.K., Wayne D., Barclay A.N., Anstee D.J.:
"The Lutheran blood group glycoprotein, another member of the
RT immunoglobulin superfamily, is widely expressed in human tissues and
RT is developmentally regulated in human liver."
RL Proc. Natl. Acad. Sci. U.S.A. 92:5496-5500(1995).
RN [2]
RP SEQUENCE OF 1-588 FROM N.A.
RX MEDLINE-95042297: Pubmed-7954395:
RA Campbell I.G., Foulkes W.D., Senger G., Trowsdale J.,
RA Garin-Chesa P., Rettig W.J.:
RT "Molecular cloning of the B-CAM cell surface glycoprotein of
RT epithelial cancers: a novel member of the immunoglobulin
RT superfamily."
RL Cancer Res. 54:5761-5765(1994).
CC -1- FUNCTION: PROBABLE RECEPTOR. MAY MEDIATE INTRACELLULAR SIGNALING.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: WIDE TISSUE DISTRIBUTION (HIGHEST IN THE

```

CC PANCREAS AND VERY LOW IN BRAIN). CLOSELY ASSOCIATED WITH THE BASAL
 CC LAYER OF CELLS IN EPITHELIA AND THE ENDOTHELIUM OF BLOOD VESSEL
 CC WALLS.
 CC -1- DEVELOPMENTAL STAGE: IS UNDER DEVELOPMENTAL CONTROL IN LIVER AND
 CC MAY ALSO BE REGULATED DURING DIFFERENTIATION IN OTHER TISSUES.
 CC UPREGULATED FOLLOWING MALIGNANT TRANSFORMATION IN SOME CELL TYPES.
 CC -1- POLYMORPHISM: LD IS RESPONSIBLE FOR THE LUTHERAN BLOOD GROUP
 CC SYSTEM.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X83425; CAA58449.1; -
 DR EMBL; X80026; CAA56327.1; -
 DR MIM; 111200; -
 DR INTERPRO; IPR003006; -
 DR PFAM; PF00047; 1g; 5.
 KW Receptor; Immunoglobulin domain; Glycoprotein; Transmembrane; Signal;
 KW Blood group antigen.
 FT SIGNAL 1 31
 FT CHAIN 32 628 LUTHERAN BLOOD GROUP GLYCOPROTEIN.
 FT DOMAIN 32 547 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 548 568 POTENTIAL.
 FT DOMAIN 569 628 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 46 132 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 165 244 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 284 344 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 377 431 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 466 529 IG-LIKE C2-TYPE DOMAIN.
 FT DISULFID 53 125 PROBABLE.
 FT DISULFID 172 237 PROBABLE.
 FT DISULFID 291 337 PROBABLE.
 FT DISULFID 384 424 PROBABLE.
 FT DISULFID 473 522 PROBABLE.
 FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 439 439 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 225 226 RL -> PC (IN REF. 2).
 FT CONFLICT 355 356 EL -> DV (IN REF. 2).
 SQ SQUENCE 628 AA; 67374 MW; C89B0A4835492B1E CRC64;

Query Match 19.3%; Score 647.5; DB 1; Length 628;
 Best Local Similarity 29.5%; Pred. No. 1.5e-35;
 Matches 202; Conservative 103; Mismatches 265; Indels 115; Gaps 27;

QY 2 GUPR-LYCAALLAACCCCPVAVGPEBAEPALVEVEGSTALLKGLSGSOGMLSH- 59
 DB 12 GAPPRLLLAVLLAA-----HPDAQAEVRLSVPEPLVEVMGKSGVILDC--TPTGTHDHX 62
 QY 60 -VWME-----SVHKEKRTLLIFRYRGOGGSGPEYRQRSLDQKGTALTALTOY 106
 DB 63 MLEMFLLDRSGARPLASAMOGSELQVTHMDTRGKSP---YOLDSGR--LVLAFA 115
 QY 107 TPQDERIFLC--OGKRRROSEYRIOLRVYKAPPEPNIQVPLGIPVNSKEPEVATCVGR 164
 DB 116 QVDEDERYVAVRAGAAGTATARLNVFAKPEATEVSPNKGTLSWMSADQIACNSR 175
 QY 165 NGPIPIQVIVYKNGRPLK--EENRKHIOSSQTV--BSSGLYTLQSLKQVLKEDKDAQF 221
 DB 176 NGNPAPRITVYNGORLEPVVENMPEGVMTSRVTEASGLISTFLYRLRKDRDASAF 235
 QY 222 YCELNRLPSGNHMK--ESREVTVPVYPTPEKY--WL--EVEPYGMKEDDREIRCLAG 276

DB 236 HCAAHYSLEGRHGRDLSPTFHLTLHYPTHEHGVFWGSPSTRAGVREGDTYOLLCRGDG 295
 QY 277 NPPPHSISKQNSTEAEETTNDN--GVLYLEPARKHSGRYECGGIDDTMTSL-LS 333
 DB 296 SPSEYTLFR---LQDEDEVLNVMLEGNLTLEGVTRQSGTYGCGRVEDYDAADVQLS 351
 QY 334 EPELLVNVSDVYRVPAPAREQEG-----SSLTLTCEANSSODLEFQVLRETFQY 385
 DB 352 KTEFLRAYATIDPLELS-----EGKVLSLPLNSSAVVNCVHGILPTPALRTKDT--P 402
 QY 386 LERGPVYLQHLDKREAGGRCVAVSPISPGINRTQLVAVAFGPPMAFKERK-----V 440
 DB 403 LGGPMPLSSLTIFEDNNGTYVCASLPTVPVLSRTQNFLLVYGSEPLKTAIEPKAGS 462
 QY 441 WVENNVNLINSCASGHPPTTISWNVNGTASRQDDQPR---VLSTLVNLYTPELLETGY 497
 DB 463 W-REGDEVLTISARCHPPKLSWSQLG--GSPAEPFPGQGVWSSSLTLKVTALSRLGI 520
 QY 498 ECTASNDLKNSTSLFLELVNLTTLTPDSNTTGLSTASPHRTANSTREKLPPEES 557
 DB 521 SCASNPNGNKRVRHFEGAVS-----PQTSQ 546
 QY 558 RGVYIVAVTICLLVLAVALGAVLYFYKKGKLPGRSGKOETLPPSRKSELVVEYKSDKL 617
 DB 547 AGVAVMAVAVSVGLLLVAVVFCVARRKG--PCQRQREKGRPP---GEGTSHSGSEQ 602
 QY 618 PREMGLQG--SSGDKRAPGDQGEK 640
 DB 603 PRQGTGLMGASGAGGSGGFGDE 627

RESULT 3
 ID C166-CHICK STANDARD; PRT; 588 AA.
 AC P42292;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CD166 ANTIGEN PRECURSOR (SCL GLYCOPROTEIN) (BEN GLYCOPROTEIN) (DM-
 DE GRASP PROTEIN) (JC7 PROTEIN).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-53.
 RC TISSUE=EMBRYO;
 RX MEDLINE=92030150; PubMed=1931049;
 RA Tanaka H., Matsui T., Agata A., Tomura M., Kubota I.,
 RA McFarland K.C., Kohr B., Lee A., Phillips H.S., Shelton D.L.;
 RT "Molecular cloning and expression of a novel adhesion molecule, SCL.",
 RL Neuron 7:535-545(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91337449; PubMed=1873027;
 RA Burns F.R., von Kannen S., Guy L., Raper J.A., Kamholz J.,
 RA Chang S.;
 RT "DM-GRASP, a novel immunoglobulin superfamily axonal surface protein
 RT that supports neurite extension.",
 RL Neuron 7:209-220(1991).
 RN [3]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-52; 87-93; 504-512 AND 569-582.
 RC TISSUE=BURSA OF FABRICIUS;
 RX MEDLINE=92302224; PubMed=1608932;
 RA Pourquie O., Cortel C., le Caer J.-P., Rossier J., le Douarin N.M.;
 RT "BEN, a surface glycoprotein of the immunoglobulin superfamily, is
 RT expressed in a variety of developing systems.",
 RL Proc. Natl. Acad. Sci. U.S.A. 89:5261-5265(1992).
 RN [4]
 RP POSSIBLE FUNCTION.
 RX MEDLINE=92211411; PubMed=1313497;
 RA Pourquie O., Hallonet M.E.R., le Douarin N.M.;
 RT "Association of BEN glycoprotein expression with climbing fiber

RT axonogenesis in the avian cerebellum.";
 RL J. Neurosci. 12:1548-1557(1992).
 CC -1- FUNCTION: HOMOPHILIC ADHESION MOLECULE ASSOCIATED WITH CLIMBING
 CC FIBER AXONOGEMESIS. SUPPORTS NEURITE EXTENSION.
 CC -1- CELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A RESTRICTED POPULATION OF AXONS.
 CC WITHIN THE SPINAL CORD IT IS LOCALIZED TO AXONS IN THE DORSAL
 CC FUNICULUS, MIDLINE FLOOR PLATE CELLS, AND MOTONEURONS. FOUND IN
 CC EPITHELIA, CENTRAL NERVOUS SYSTEM, AND HEMOPOIETIC CELLS. EACH
 CC DEVELOPING SYSTEM CARRIES DIFFERENTLY GLYCOSYLATED MOLECULAR FORMS
 CC OF BEN.
 CC -1- DEVELOPMENTAL STAGE: WIDELY EXPRESSED DURING EMBRYONIC
 CC DEVELOPMENT.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S63276; AAB20170.1; -
 DR EMBL: M76678; AAA48602.1; -
 DR EMBL: X64301; CAA45579.1; -
 DR HSSP: Q13740; IKCC.
 DR INTERPRO: IPR000495; -
 DR INTERPRO: IPR003006; -
 DR PIRAM: PF00047; 19; 5.
 DR PROSITE: PS00280; IG_MHC; FALSE_NEG.
 KM Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;
 FT Signal. 1 33
 FT CHAIN 34 588
 FT DOMAIN 34 532
 FT TRANSMEM 553 553
 FT DOMAIN 554 588
 FT DOMAIN 42 126
 FT DOMAIN 156 233
 FT DOMAIN 269 326
 FT DOMAIN 352 404
 FT DOMAIN 433 497
 FT DISULFID 49 119
 FT DISULFID 163 226
 FT DISULFID 276 319
 FT DISULFID 359 397
 FT DISULFID 440 490
 FT CARBOHYD 101 101
 FT CARBOHYD 173 173
 FT CARBOHYD 199 199
 FT CARBOHYD 271 271
 FT CARBOHYD 312 312
 FT CARBOHYD 366 366
 FT CARBOHYD 462 462
 FT CARBOHYD 485 485
 FT CARBOHYD 504 504
 FT CONFLICT 2 10
 FT CONFLICT 25 25
 FT CONFLICT 112 113
 FT CONFLICT 329 329
 FT CONFLICT 401 402
 FT SEQUENCE 588 AA; 65726 MW; 2A28612D0164531E CRC64;

Query Match 15.28; Score 509.5; DB 1; Length 588;
 Best Local Similarity 26.48; Pred. No. 1.8e-26;
 Matches 166; Conservative 117; Mismatches 254; Indels 91; Gaps 30;

QY 6 LVCAFLAACCCCPVGVPEAGPAPELVGVGSGTALCKQLSOGNLSHVDW-FS 64
 DB 18 LLC--LLTAAALCPALGL-----YTVNAVYGDITTPCRLVEYPDG-LMEGKKKYE 65

QY VHKERTLI-FR--VRQGGQSGPEYEQRLSDRATLALTVQTPDERIFCQ--VGK 119
 DB 66 MENSPPVIARSSSKKVVQDDPDYDKRLSIE-NYTLISIKARISDEKRFVCMYTE 124
 QY 120 RPRSGEYRIQRLVYKAPKEPNI-QVNDLIPVNSKEPEEYAVCVGNGYPIPOVIYKN 177
 DB 125 DVSEEPVY-VKVEKQPSQPELHQADEL-----ETEKLMKGECVVDSDYDEGNVYKN 179
 QY 178 GRPLK--EKKRVHIIQSSQYVSSSLYTLQSLIAQLVKEKXDAQFCELYRLPQGNHM 235
 DB 180 GRVLQPVVEEVVYINLRKEN-RSTGLFTYTSLLQYMPTEKANKFVCIYTHGSPQKT 238
 QY 236 KSRREVTVPEVPEKRWLEW-EPYGMLESGDRVEIRCLADGNPPH---FSISKQNPST 291
 DB 239 IQSEPVVDVHYPTPKVYIRVLSQSSITKEDDNTYTKSCGNGNPPQPELFYI----- 291
 QY 292 REAEETNDNGVLEVPAREKHSRGECGLDPTMISLSEFOELLYVSVYRSPA 351
 DB 292 -PGETEGIRSSDPTVYMTVRRNATGEYKCSLIDKSM-----DATYIVHYL-DLQLTPS 344
 QY 352 AP-EROEGSSILLTNGEARSODLEFQWLRFEFGVLEGPVLOLHDLKREAGGGRCVAS 410
 DB 345 GEVTKQIGBALPVSCYTISSSRNATYFWIKDNR--MKTSP--SFSSIQYQAGNYICETT 400
 QY 411 VPSIFGLNRTQLVNAIFGPPMAFERKRVYKENVYLNLSCEASGHPRTISMNVNGTA 470
 DB 401 LQEVGLKRRKTLKLIVGKP--QIKMTKNTNMKSTVICHVEGFPAVQWTVVGS 458
 QY 471 S--EODQDQRLSTLN--VLATPELLETGYECVTSNOLGNTSLFLVNLNTLPDS 526
 DB 459 SLINTEKRYVNGKFSKIIIAPEENVY-LTCLAEHL-----E 497
 QY 527 NNTTGLSTSTAS-----PHTRANSTTERKLPEPESRGVYIVAVICLVAVGAVLY 580
 DB 498 KTVTSLSNVAISIPYDEDEDR-NDNSEK-----VNDQAKLIVGLVGLLVALVAVGVY 552
 QY 581 FLY-KKGLPCRRSGKQETLTPPSRKE 607
 DB 553 WLYVKKSTASKVHVDKDLGNIEKNKLE 580

RESULT 4
 C166_HUMAN STANDARD; PRT; 583 AA.
 AC Q13740; O60892;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CD166 ANTIGEN PRECURSOR (ACTIVATED LEUKOCYTE-CELL ADHESION MOLECULE)
 GN ALCAM.
 OS Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95279947; PubMed=7760007;
 RA Bowen M.-C., Margard H., Neubauer M., Pesano J.M., Francke U.,
 RA Wang W.-C.,
 RA Haynes B.F., Aruffo A.;
 RT "Cloning, mapping, and characterization of activated leukocyte-cell
 RT adhesion molecule (ALCAM), a CD6 ligand.";
 RL J. Exp. Med. 181:2213-2220(1995).
 RN [2]
 RP SEQUENCE OF 2-583 FROM N.A.
 RX MEDLINE=98161527; PubMed=9502422;
 RA Degen W.G., van Kempen L.C., Gijzen E.G., van Groningen J.J.,
 RA van Kooyk Y., Bloemers H.P., Smit G.W.,
 RT "MEMD, a new cell adhesion molecule in metastasizing human melanoma
 RT cell lines, is identical to ALCAM (activated leukocyte cell adhesion
 RT molecule).";
 RL Am. J. Pathol. 152:805-813(1998).

RA Kanki J.P., Chang S., Kuwada J.Y.:
 RT "The molecular cloning and characterization of potential chick
 RT DM-GRASP homologs in zebrafish and mouse."
 RL J. Neurobiol. 25:831-845(1994).
 CC -1- FUNCTION: CELL ADHESION MOLECULE THAT BINDS TO CD6. INVOLVED IN
 CC NEURITE EXTENSION BY NEURONS VIA HETEROPHILIC AND HOMOPHILIC
 CC INTERACTIONS. MAY PLAY A ROLE IN THE BINDING OF T AND B CELLS TO
 CC ACTIVATED LEUKOCYTES. AS WELL AS IN INTERACTIONS BETWEEN CELLS OF
 CC THE NERVOUS SYSTEM.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U95030; AAC06342.1; -
 CC MGD: L25274; AAA37528.1; -
 CC DR INTERPRO: IPR000485; -
 CC DR INTERPRO: IPR003006; -
 CC DR PFM: PFO0047; 19; 5.
 CC DR PROSITE: PS00290; IG-MHC; FALSE_NEG.
 CC KW Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;
 KW Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 583
 FT DOMAIN 28 527
 FT TRANSMEM 528 549
 FT DOMAIN 550 583
 FT DOMAIN 263 321
 FT DOMAIN 347 399
 FT DOMAIN 428 492
 FT DISULFID 43 113
 FT DISULFID 157 220
 FT DISULFID 270 313
 FT DISULFID 354 392
 FT DISULFID 435 485
 FT CARBOHYD 95 95
 FT CARBOHYD 167 167
 FT CARBOHYD 265 265
 FT CARBOHYD 306 306
 FT CARBOHYD 361 361
 FT CARBOHYD 457 457
 FT CARBOHYD 480 480
 FT CARBOHYD 499 499
 FT CONFLICT 227 232
 FT CONFLICT 454 454
 FT SEQUENCE 583 AA; 65161 MW; E7BAF8BCA8F9489 CRC64;
 Query Match 14.58; Score 488.5; DB 1; Length 583;
 Best Local Similarity 24.58; Prod. No. 4.4e-25;
 Matches 152; Conservative 122; Mismatches 274; Indels 73; Gaps 25;

DB 176 VLOVEGEVALLFEKLEIDPGQLTAVTSSLEKTTRESDIQMPFTCSVTVYGGPSGQRTYS 235
 QY 239 REVTVPVFPYETKWEVLEP-EPVGMALKEGDRVEIRCLADGNPPH---FSTSKONPSTREA 294
 DB 236 EQEIFDIYPTPEQYTVLPPKNAIKEGDNTLLOCLONGNPPEEFMYLPQGPGRIRSS 295
 QY 295 EEFITNDNGVLELPARKHSGRECOGLDPTITSLSEPOELVNVYSDVRSVPAAP- 353
 DB 296 NPTLTLD-----VRNATGDKCSLIDKRMAAST---ITVHYL-DLSINPRGEV 342
 QY 354 EROGSSSLTTCSEASSODLEFOMLREETQVLERGVLOLHDIKRAGGCIYCVASVPS 413
 DB 343 TKQIGDITLPVSCSTIASRNATVVMKDNIR--LRSSP--SFSSHYODAGNVCETALQE 398
 QY 414 IPLGNKQVLVAVLFGPPWMAFERKRVKVMKNMVLNLSCEASGPRPTISMNVTGAS-- 471
 DB 399 VEGLKRRRESLTLIVEGKP--QIKMTKTPDSGLSKTTICVHEGPKPAIMHTTIGSGSVI 456
 QY 472 -EODDDPQ---RVLSLAVLVTPLELTGVECTASNDLGKNTSILFELVNLTPPDSN 527
 DB 457 NQTESPIYNGRYYS--KIISPEENVY-LCTAENQLEFVNSLVNSAISL---PE-- 507
 QY 528 TTTGLSTSPHTRANSTSTERKLPPEPSRGVYIVAVICVLVAVLGAVALFTLY-KKG 586
 DB 508 -----HDEADDISDENR-EKVNDQAKLIVIGVLLALAVAGVYVWLYMKKS 554
 QY 587 KLPGRSGKOETITPPSRKE 607
 DB 555 KTSKRVNKNDLGNMENKKE 575

RESULT 6
 C166 CARAU STANDARD: PRT: 555 AA.
 ID C166 CARAU
 AC 090304; 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CD166 ANTIGEN HOMOLOG PRECURSOR (NEUROLIN) (DM-GRASP HOMOLOG).
 OS Carassius auratus (Goldfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Cyprininae; Carassius.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RETINA;
 RC TISSUE=RETINA;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 30-552 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=RETINA;
 RC MEDLINE=94299040; PubMed=8026643;
 RA Laessing U., Giordano S., Stecher B., Lottspeich F., Stuermer C.A.O.;
 RT "Molecular characterization of fish neurolin: a growth-associated
 RT cell surface protein and member of the immunoglobulin superfamily in
 RT the fish retinotectal system with similarities to chick protein
 RT DM-GRASP/SC-1/BEN.";
 RL Differentiation 56:21-29(1994).
 CC -1- FUNCTION: CELL ADHESION MOLECULE. INVOLVED IN NEURITE EXTENSION BY
 CC NEURONS VIA HETEROPHILIC AND HOMOPHILIC INTERACTIONS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: PRESENT ON ALL RETINAL GANGLION CELLS (RGCs)
 CC AND THEIR AXONS (IN EMBRYO). ABSENT FROM MATURE AXONS ALONG MOST
 CC OF THEIR LENGTH, BUT IS PRESENT ON NEW AND GROWING AXONS DERIVED
 CC FROM THE RGCs AT THE RETINAL MARGIN. REMAINS ON ADULT RGCs ONLY AT
 CC CELL-CELL CONTACT SITES AND IS CONTINUOUSLY FOUND IN THE RETINAL
 CC AXON TERMINAL ARBOR LAYERS OF THE ADULT TECTUM.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: L25056; AAC38015.2; -

DR HSSP: Q13740; IKJC.

DR INTERPRO: IPR000495; -

DR INTERPRO: IPR003006; -

DR PFAM: PF00047; 1g; 3.

DR PROSITE: PS00290; IG_MHC; 1.

KW Signal, Developmental protein; Cell adhesion; Immunoglobulin domain;
 KM Glycoprotein; Transmembrane.

FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 555 CD166 ANTIGEN HOMOLOG.
 FT DOMAIN 23 499 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 500 520 POTENTIAL.
 FT DOMAIN 521 555 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 31 117 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 147 224 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 256 313 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 340 392 IG-LIKE C2-TYPE DOMAIN.
 FT DOMAIN 419 477 IG-LIKE C2-TYPE DOMAIN.
 FT DISULFID 38 110 POTENTIAL.
 FT DISULFID 154 217 POTENTIAL.
 FT DISULFID 263 306 POTENTIAL.
 FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 555 AA; 60371 MW; 5A4AB014F00BF68 CRC64;

Query Match 13.88; Score 465.5; DB 1; Length 555;
 Best Local Similarity 26.58; Pred. No. 1.4e-23;

Matches 167; Conservative 111; Mismatches 234; Indels 119; Gaps 27;

QY 6 LVCAFLAAC-----CCCPVAVGPGEAQPAPELVEVGSSTALLKC--GLSOSQGLSH 59
 Db 7 LIGAFIAAAYRRPSCGYVIGLGE-----TIVPNDCTKKPDG-LIF 50
 QY 60 VDMFSVHKERTLLIFRVKGGQ-----SEGEYEQRLSLDRGNTIALTVTPQDERIFL 115
 Db 51 TKMKYVDDGSPDLLVKQKQKDEATVSATDGYKSRVSA--ANSLLIANGSLADQVFT 109
 QY 116 CQ-GKRRSQEYRIQLVYKAPPEPNIQVPLGIPVNSKEPE-----EVATCVGRNGYPT 169
 Db 110 CMVVSFTNLEYSVEYVHKRPAPVTK-----NNAKELENGKLTQLGECVVENANPP 162
 QY 170 PVIWYKNGRPLKEKRNKRVHIOSSQVES--SGLYTLOSILKAOLVKRDXAOFCYELNTR 228
 Db 163 ADLWKNKNTLVDDGKTIITITITDKITGLTSSSTRLOIYARKKDEVSQFTC----- 217
 QY 229 LPSGNM-----KSRREYTVPEYVPTKEKWLVEVEVGMALKGDVEIKCLADGNPP--H 281
 Db 218 --TAKHWMPDQVSEPEFPIHYPTKEVLSQVYOSPIRGEDEVTLKCAQDGNPPPTSEN 275
 QY 282 FSTSKQNPSTRAEETNTNGVILEPARKHSGRECOGLDLDTIVLISLSEPEELVN 341
 Db 276 FNI-----KGRKVVTVDKDYVTLGLTVRADSGIKYKCSLNDVWES-----TQFVVS 323
 QY 342 YVSDVAVSPAPE--RQEGSLTLTCEASSQDLEFQWLAREFTGOVERGVQLHDKRE 400
 Db 324 FL-DVSLITPTGKVLKNGENLIVSLDKMNSSEAKVWTTD--NRKLDKL--DFSKLTIS 378
 QY 401 AGGCIKVAVSPISPGINPTQLVNVAIFGPPWMA--FKERKVVAKENMVLNLSCEASGHP 458
 Db 379 DAGLYVCDV--SIEGIKRSLSELTVEGIPKITSLSLKHSSDGKHKV--LTCEAEGSP 432
 QY 459 RPTISMVNGTASEQDDPDRVLTSLNVLVTPLELTYGECTASNDLGKNTSLLELVN 518

Db 433 KPDQWVNGENDESVNNGKATYKLTIVSKULT---VSLVTNKLGETET----- 483
 QY 519 LTTITPDSTNTTGLISTASPHRANSTSTERKLPPESKGVIVAVICILVAVLAV 578
 Db 484 -----SVFSQKNEDGTE-----QAKYIVGIWGLLVAAALVGL 516
 QY 579 LYFLY-KKGLPKRRSGKQEL-TLPSRKSE 607
 Db 517 IYWIYIKTRQSGWKGTGEKAGESESKLE 547
 RESULT 7
 ID C166_BRARE STANDARD; PRT; 564 AA.
 AC 090460; 090480;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CD166 ANTIGEN HOMOLOG PRECURSOR (NEUTROLIN) (DM-GRASP HOMOLOG).
 GN CD166.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 OC Cypriniformes; Cyprinidae; Rasbora; Danio.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94376084; PubMed=8089660;
 RA Ranki J.P., Chang S., Kuwada J.Y.;
 RT "The molecular cloning and characterization of potential chick
 RT DM-GRASP homologs in zebrafish and mouse."
 RL J. Neurobiol. 25:831-845(1994).
 RN [2]
 RP SEQUENCE OF 398-561 FROM N.A.
 RX MEDLINE=94299040; PubMed=8026643;
 RA Laessing U., Giordano S., Stecher B., Lottspeich F., Stuermer C.A.O.;
 RT "Molecular characterization of fish neurolin: a growth-associated
 RT cell surface protein and member of the immunoglobulin superfamily in
 RT the fish retinotectal system with similarities to chick protein
 RT DM-GRASP/SC-1/BEN."
 RL Differentiation 56:21-29(1994).
 CC -1- FUNCTION: CELL ADHESION MOLECULE. INVOLVED IN NEURITE EXTENSION BY
 CC NEURONS VIA HETEROFILIC AND HOMOPHILIC INTERACTIONS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC 3 C2-LIKE AND 2 V-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

FT DOMAIN 340 389 IG-LIKE C2-TYPE DOMAIN
FT DOMAIN 419 477 IG-LIKE C2-TYPE DOMAIN
FT DISULFID 38 110 POTENTIAL
FT DISULFID 154 217 POTENTIAL
FT DISULFID 263 306 POTENTIAL
FT DISULFID 426 470 POTENTIAL
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CONFLICT 488 497 LEEDKPKPG -> R (IN REF. 2)
SQ SEQUENCE 564 AA: 61273 MW: ADA7C293A607929 CRC64;

Query Match 13.0%; Score 438; DB 1; Length 564;
Best Local Similarity 26.0%; Pred. No. 9.2e-22;
Matches 161; Conservative 112; Mismatches 254; Indels 92; Gaps 26;

7 VCAFLAACCCPRVAGPEAQPAPAEVEVSGTALLKGLSOSQNLISHVDFSVH 66
12 IAAALFAPGSCLPVIGLYGET-----IEVPCNN-----GNNKPDG-LITTKKKYAK 57
67 KEKRTLLFRVROGOGQ---SERGEYEQRLSLDRGATLALQVTPDERIFLQ-GRKP 121
58 DDGSPGDLLIKQAOQKDDPTVSAMDGKTRVISA-ANSSLLIAOGSLTDQHFVTCMVVSSP 116
122 RSQERYQLQVYKARPEPNIOVPLGIPVNSKEPEEATCGVNGNGPIQVYWKNGRPI 181
117 NLEEFSEVAVHKKPAPVIR-KNKVELENGK-LTQEGEVESANPADLIMKKNQAL 174
182 KEENRHHIOSQTVES-SGLYTLQSLILKAQLKEDKDAQFYCELYNLRPSGNHRESRE 240
175 VDQCKTIIITSDVTKDPVTGLSTSSLSLQYTKAREDAVASQFTC-----VAKHVTPGNQ 227
241 VTPV-----VYPIEKVWLEVEPNGLKEGDRVEIRCLADGNPP--HESISKQNSTRE 293
228 VSTDFEQIRPYREKVSILQVVSQSPIREGDDVTLCKQADBNPPTSFNFI-----K 279
294 AEETTDNCGVLELPARKHESGRYEGQGLDIDTMIISLSEPOBLVANYVDVSPAP 353
280 GKRYVTYDKDYYTLTGVTNRADSGYKCSLNDNDVME--TQIVVYSFL-DASLTPYTK 334
354 E-ROEGSSLTITCEAESSODLEFQWLRETFQVLEGRPVQLDLHDKRACGGYRCVASVP 412
335 VLKRLGELVLYSLEKNASSEKVTWTKD--NRKLDKLP--DESQLRVSADGLVYCDV--- 387
413 SIPLNRTQLVNVAIFGPPMA--FKERKVVAKENMVLNLSCEASGHPRTISNVNGTA 470
388 SIEGIKHSFSELTVEGGPRITGLTKHRSNDGKHV--LTCEAEGSPKPEVOMSVNGTD 444
471 SEQQDDQORVLTNLVLTPELTGVECTASNDLGKNTSILFLELVNLTLTPDSWTTT 530
445 DETSYVNGKATYKLTVPVSKLT---VSCLVTKLGGDT-----KDISSVSLPEEDKPKR 496
531 GLSTSTASPHTRANSTERKLPEPESRGVIVAVIICILAVLAVGLVFLY-KKCKLP 589
497 G-----KNEDGADQAKVIVGVVGLFLAALVGLIWLXYIKTKTRQG 537
590 CRSGSKQEI-TLPPSKSE 607
538 SMTGKETGTSEESKLE 556

RESULT 8
ICR_DROME STANDARD; PRT; 764 AA.
AC Q08180;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IRREGULAR CHIASM C-ROUGHEST PROTEIN PRECURSOR (IRREC PROTEIN).
GN RST.
OS Drosophila melanogaster (Fruit fly).

```

```

OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta;
OC Pterygota: Neoptera: Endopterygota: Diptera: Brachyera: Muscomorpha;
OC Ephydroidea: Drosophilidae: Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:94102535; PubMed:7503814;
RA Ramos R.G., Iglol G.L., Lichte B., Baumann U., Maier D.,
RA Schneider T., Brandstaetter J.H., Froehlich K.-F.;
RT "The irregular chiasm C-roughest locus of Drosophila, which affects
RT axonal projections and programmed cell death, encodes a novel
RT immunoglobulin-like protein."
RL Genes Dev. 7:2533-2547(1993).
CC -! FUNCTION: REQUIRED FOR CORRECT AXONAL PATHWAY FORMATION IN
CC THE OPTIC LOBE AND FOR PROGRAMMED CELL DEATH IN THE DEVELOPING
CC RETINA.
CC -! SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -! TISSUE SPECIFICITY: POSTEMBRYONIC EXPRESSION IS STRONG IN THE
CC DEVELOPING OPTIC LOBE AND IN THE EYE IMAGINAL DISC.
CC -! DEVELOPMENTAL STAGE: STRONGLY EXPRESSED IN EMBRYOS. ALSO FOUND
CC IN LATE LARVAL AND PUPAL STAGES.
CC -! SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 221641; CAN79756.1; -
CC EMBL: L11040; AAA16632.1; -
CC PIR: A49448; A49448.
CC PIR: S34129; S34129.
CC HSSP: P56276; ITRK.
CC FLVBASE: FBgn003285; rst.
CC INTERPRO: IPR003006; -.
CC PFAM: PF00047; 19; 5.
CC KW Transmembrane; Immunoglobulin domain; glycoprotein; Signal;
CC cell adhesion.
CC FT SIGNAL 1 19
CC FT CHAIN 20 764
CC FT DOMAIN 20 533
CC FT TRANSMEM 534 536
CC FT DOMAIN 557 764
CC FT DOMAIN 32 115
CC FT DOMAIN 137 221
CC FT DOMAIN 245 261
CC FT DOMAIN 265 332
CC FT DOMAIN 353 414
CC FT DOMAIN 432 515
CC FT DOMAIN 637 660
CC FT DOMAIN 211 211
CC FT CARBOHYD 211 211
CC FT CARBOHYD 393 393
CC FT CARBOHYD 393 393
CC FT CARBOHYD 400 400
CC FT CARBOHYD 507 507
CC FT SEQUENCE 764 AA: 82947 MW: 262225D2B2A1C181 CRC64;

Query Match 8.0%; Score 270; DB 1; Length 764;
Best Local Similarity 20.7%; Pred. No. 1.8e-10;
Matches 137; Conservative 89; Mismatches 277; Indels 158; Gaps 25;

40 VGSTALIKGLSOSQNLSHV-DMFVHKEKRLTLFRVROGOGSEPEYDQRLSLDRG 98
41 VGARVTLPCVINKQGLTQTKDQDFGLGTSRDLGSEFRAMYGSDDEGY----- 90
99 ATLALQVTPDERIFLQCK-----RPSQERYQLQVYKARPEPNIOVPLGIPVNSK 153
91 SLDIYPMDDARQCVSPGEGQPAIRSFAGLTVLPPEAKITQGGDIYATADR 149
154 EPEEVATCGVNGNGPIQVYWKNGR-----FLKEKKNVHIIQSSQTVESSGL 201

```

```

Db      150 KVE--IECVSGKPAEETWIDGCVLTIDNIEYVPLEDQR----- 192
Oy      202 YLQSLTILKAQVLKEDKQDFCELYRLPSGNHMKSESEVTVPVFPTKYLWEVEPYGM 261
Db      193 FPAKSVLRLTPKKEHHNNTFSCAO--NTADRTYRSAKIREVYKAK--VKVNWGS 246
Oy      262 LKEG-----DVEIRCLAGCNPP--HFSISKQNSTREAE 295
Db      247 LPGAAGSVGAGGSGVMSKTSRIVERHSGVRLCEKRAANSDVRYRMFINDEPLIGQK 306
Oy      236 EETINDNGVLEPARKBHSRGRECOGLDPTMISLSEPOELVNVYSDVRSPAPER 355
Db      307 TE-----MVIKNVYRKFDHAIYK--EYQNSVKSJSDSETLIDISTVSPRQRQSMEA 357
Oy      356 QEGSSLLTLCESQDLEFOMLRETCQVLERGPVQLDHLKRBAGGVCVAVSPSIP 415
Db      358 DVGSVVSLTCEVDSNPDEIYWIQHPSDRVVGTSTLTF--SVSNETAGRYCKANVPGYA 416
Oy      416 GLNRTQLVNVAIFGPPMAFKERKVVYKENVVLNLSGASHPRPT--ISMVNG--TASE 472
Db      417 EISNAYVYK--GSPAIG--SQRTQYGLVGTARIKCPASSVPARRHVSMTENGQELTSE 473
Oy      473 QDQD-----POVLTSLNVLTPELLETGVECTASNDLCKNTSLIFELVMTLTLP 524
Db      474 SGHDYSILVDVAPGVGKSTLIRDSQAYHYGKYNTVNDYGNVVAELQAKKSVL-- 531
Oy      525 DSNITGISTASTHPRANSTERKLPESRGVIVAVTICVLAVGAVLYFLYK 584
Db      532 -----LMTVVG-----ISVAVELVLTLL--VVVY---- 555
Oy      585 KGLPCRRSGKOEITLPPSRKSELVEYKSDKLPEMGLLOGSGDDRAPDQGEKIDL 644
Db      556 --IKCKKRTK--LPP-----ADVISEHQITKNKGSCHLEDRDRTSNYSDL 597
Oy      645 R 645
Db      598 K 598

RESULT 9
PTK7_CHICK STANDARD; PRT; 1051 AA.
AC Q91046;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TYROSINE-PROTEIN KINASE-LIKE 7 PRECURSOR (KINASE LIKE PROTEIN).
GN PTK7 OR KLG.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYONIC BRAIN;
RA MEDLINE=91271300; PubMed=1711213;
RX Chou Y.-H., Hayman M.J.;
RT "Characterization of a member of the immunoglobulin gene superfamily
RT that possibly represents an additional class of growth factor
RT receptor."
RL Proc. Natl. Acad. Sci. U.S.A. 88:4897-4901(1991).
CC CC -1- FUNCTION: MAY FUNCTION AS A CELL ADHESION MOLECULE. LACKS PROBABLY
CC CC THE CATALYTIC ACTIVITY OF TYROSINE KINASE.
CC CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC CC -1- TISSUE SPECIFICITY: EXPRESSED IN BONE MARROW, SPLEEN, BURSAS,
CC CC THYMUS AND BRAIN. WEAKLY EXPRESSED IN FIBROBLASTS. ALSO EXPRESSED
CC CC IN EMBRYONIC LIVER.
CC CC -1- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
CC CC PROTEIN KINASES. TRK-TYPE SUBFAMILY.
CC CC -----
CC CC This SWISS-Prot entry is copyright. It is produced through a collaboration

```

```

CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC CC or send an email to license@isb-sdb.ch).
CC CC -----
CC DR EMBL; M63437; AAA48933.1; -.
CC DR HSSP; P00523; PTK.
CC DR INTERPRO; IPR000719; -.
CC DR INTERPRO; IPR001245; -.
CC DR INTERPRO; IPR003006; -.
CC DR PRAM; PF00047; 19; 7.
CC DR PRAM; PF00069; pkinase; 1.
CC DR PRINTS; PR00109; TYRKINASE.
CC DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
CC DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC KW Receptor; Transmembrane; Signal; Glycoprotein; Cell adhesion;
CC KW Immunoglobulin domain; Repeat.
CC FT CHAIN 1 22
CC FT SIGNAL 1 22
CC FT TRANSMEM 23 1051
CC FT TRANSMEM 23 685
CC FT TRANSMEM 686 706
CC FT DOMAIN 707 1051
CC FT DOMAIN 130 95
CC FT DOMAIN 130 194
CC FT DOMAIN 227 289
CC FT DOMAIN 319 379
CC FT DOMAIN 407 469
CC FT DOMAIN 498 558
CC FT DOMAIN 587 652
CC FT DOMAIN 777 1048
CC FT DISULFID 40 88
CC FT DISULFID 137 187
CC FT DISULFID 234 282
CC FT DISULFID 326 372
CC FT DISULFID 414 462
CC FT DISULFID 505 551
CC FT DISULFID 594 645
CC FT CARBOHYD 103 103
CC FT CARBOHYD 202 202
CC FT CARBOHYD 255 255
CC FT CARBOHYD 264 264
CC FT CARBOHYD 444 444
CC FT CARBOHYD 548 548
CC FT CARBOHYD 627 627
CC SQ SEQUENCE 1051 AA; 116366 MW; 1752442AEA4CB702 CRC64;

Query Match 7.7%; Score 259.5; DB 1; Length 1051;
Best Local Similarity 20.1%; Pred. No. 1.4e-09;
Matches 149; Conservative 89; Mismatches 263; Indels 241; Gaps 31;

Oy 29 EQPAELVEVEYSTALLKCGLSQSGNLSHYD-----WF-----SVHK 67
Db 119 KQPA-SAAIOPSSIVLRC-----HIDGHPRTWQFRGAPLPGRGYTVSS 167
Oy 68 EKRTLIFRVGOGGSEPEY-----DORTSLDRGATVLTQ-----V 106
Db 168 KERTITLR--GAGPDDNGLYYCSARPAVGSVCSDQNFNTIIDESPQAVVPEDLIV 224
Oy 107 TPQDERIFLCO--GKRPSQOEY-----RIQLR--YK-- 134
Db 225 TNERAMFDCQAAVPPPTQEWLFEDSPITNSKTTVANGSLILITQKARSTGVYKCIQ 284
Oy 135 -----APEEPNIQVNPDLGIPYNSKEPEVATCVARNGPIPOV 172
Db 285 HGQKKAIVLKATLRRLAIEEMAPSPKVLNANGHR-----SCAPRPVPIPOV 335
Oy 173 IYKKN-----GRLKEKNRNVHIQSSQVYESSGLTYLQSLKAQVLKEDKQDFYCEL 225
Db 336 WMRNOERVPTAGRYQEAEDLVF--TSITBADAGIYCHAAKNA----- 378

```

```

OY 226 NVRLESGNHHKRESREYTVFVPEKVMLEVEPVGMLEKEDRVEIKLADGNPPPHFSIS 285
DB 379 -----GKKOELITV-----ATVPKVMKPKDSOLESKCYLHLSKASLKPRVWY 427
OY 286 KNPSTREAEEETNDNGVLYLEPARKHSGRYECGLDLDIMISLSEPOELVNY 342
DB 428 RNCVJSLESDSRFISNGTLRINNVEYDGTMYK-----VSPASISIEGYAVH 477
OY 343 --VSDRVSPAPBERQ---EGSSLITITCAESSQDLEFQMLRE---IGCVLERGPVQ 393
DB 478 HYLEKIKFTPPPOPLQCEMEFNKEVYSCSATGREKPTIQTWTDGSSLPSSHRAGLIS 537
OY 394 LHLKREAGGCVGCVASVSPISPLNRTQLVNVAIFGPPMAFK--ERKVVVENMVLNIS 451
DB 538 FHVYSNDSNGNYTCIASNSNGEIRATVOLVAVY---VTFRLPEEPTTVQGHAMFQ 593
OY 452 CEASGHPRTISNVNGTASEODOPQVYLSTLVLTPELLETVECTASNDLGNISI 511
DB 594 CQAGDPVPHIQMK-----GKDKILDPKILPQIIMPNSLV---LYDVTEDSGKYTCI 646
OY 512 -----LFLVNLTLTPDSNTTGLSTASPHTRANSSTERKLPPEPESRGV 560
DB 647 AGNSCNKIHREAFLYVVDKPAEEDG-----PSSHTPYKMTQT-----IGL 688
OY 561 VIVAVICILVAVLGAVALFYKKGKLPGRSGKQIEITLPSRKSSEL----- 608
DB 689 SVGAAYITII--VLG--LMFYCK-----RRKAKRLKHREGEPEMECNGTILONG 739
OY 609 --VVEVSKDLPEEMGLQSS 628
DB 740 OTTAEIQEVALTNLIGSSSGAS 761

RESULT 10
NCAL_XENLA STANDARD: PRT: 1088 AA.
ID NCAL_XENLA
AC P16170:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE 1, LARGE ISOFORM PRECURSOR (N-CAM 180)
[CONTAINS: N-CAM 140].
GN NCAM1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodidae; Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90098871; PubMed=2481269;
RA Krieg P.A., Sakaguchi D.S., Kintner C.R.;
RT "Primary structure and developmental expression of a large
RT cytoplasmic domain form of xenopus laevis neural cell adhesion
RT molecule (NCAM).";
RL Nucleic Acids Res. 17:10321-10335(1989).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEURON AND IN PRESUMPTIVE NEURAL
CC TISSUE.
CC -1- DEVELOPMENTAL STAGE: THE MRNA ENCODING THIS LD-NCAM IS THE MAJOR
CC TRANSCRIPT PRESENT IN BOTH MATERNAL RNA AND IN THE EMBRYO DURING
CC EARLY NEURAL DEVELOPMENT.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL: M25696; AAA49909.1;
DR PIR: S09600; IXLNL.
DR HSSP: P56276; IFLK.
DR INTERPRO: IPR001777;
DR INTERPRO: IPR003006;
DR PFAM: PF00041; fn3; 2.
DR PFAM: PF00047; 1g; 5.
DR Cell adhesion; Glycoprotein; Transmembrane; Repeat; Brain;
KW Immunoglobulin domain; Alternative splicing; Signal.
FT SIGNAL 1
FT CHAIN 20 1088
FT DOMAIN 20 705
FT TRANSMEM 706 723
FT DOMAIN 724 1088
FT DOMAIN 34 100
FT DOMAIN 129 193
FT DOMAIN 225 289
FT DOMAIN 316 386
FT DOMAIN 413 480
FT DOMAIN 512 589
FT DOMAIN 618 686
FT DOMAIN 149 153
FT DOMAIN 158 162
FT DISULFD 41 93
FT DISULFD 136 186
FT DISULFD 232 282
FT DISULFD 323 379
FT DISULFD 420 473
FT CARBOHYD 82 82
FT CARBOHYD 219 219
FT CARBOHYD 310 310
FT CARBOHYD 341 341
FT CARBOHYD 417 417
FT CARBOHYD 443 443
FT CARBOHYD 472 472
FT VARSPIC 804 1049
SQ SEQUENCE 1088 AA; 117778 MW; 62738B5B03F3E83 CRC64;

Query Match 7.6%; Score 256; DB 1; Length 1088;
Best Local Similarity 18.1%; Pred. No. 2.5e-09;
Matches 155; Conservative 100; Mismatches 266; Indels 336; Gaps 35;

OY 33 PELVEVEVGTALNKGSLQSGNSLHVDMEFVHKEKTLIFRVAGQGQSPREYQRL 92
DB 26 PDGETSLSESKFFLC--QVSGEATDISWYSPTEKLYT-----QQOI 66
OY 93 SL---ODRGATLALNQTVPQDERIFLC-----QGR-----RPRSGEY 126
DB 67 SVVRSDYITLTIYASSQDAGIYCVASNEAGESECTVNLKTYOKLTFKNAPTPOEF 126
OY 127 R-----
DB 127 KEGEDAVIICDVSSIPSIIITWRHKGKDVIFKKDVFVYLVANNYQIKRKTDEGTRC 186
OY 139 -----
DB 187 EGRILARGELNKIDIOYIYVNPETIAROLRVNANANMAESVLSGDDAGFDPETSLMK 246
OY 177 NGRPLKEERNRVHIQSSQTVESGLYTLQSLKALQYKEDKDAQFYCELNTRLPSCNHMK 236
DB 247 KGEPIEDGEEKISFND-----QSEMTIHVKEKDAEVEYSCIAN-----NAG 289
OY 237 ESREVEVVPVFPPEKVMLEVEPVGMLEKEDRVEIKLADGNPPPHFSISKONPSTREAEE 296
DB 290 EAEATLILKAVYAPKITYVENKTAIV--ELDETLTLTCEASGDPIP--STWRTAVRNISE 345
OY 297 ETTNDNGVLYLEPARK-----HSGRYEC-----QGLDLDIMISLSEPOELVNY 341

```



```

Db 346 ATTLGDHIVKEHRLMSALTLDKIDYTDAGEYPCIASNPICVMOAMY-----FEVQ 397
QY 342 YSDVRVSPAPEROESSSTLTFCEAESSODLEFOMLREFTGOVLE-----RGVY 391
Db 398 YAKIR-GRPVYVYTWECNPVNTICEVFAHRAAVTWFRD--GOLPSSNFSNIKISGPT 454
QY 392 ---LQHLDLKREAGGRCYV-----ASVPSIPLGNLTOLVN---VAIFG 429
Db 455 SSSLEVPDSENDPGNNNCNTAINTIGHESEFLLVQADPPSSPAIKKVEYSTWAVLD 514
QY 430 P-----WMAFKKKVWK-----ENM--VLNLSCEAS----- 455
Db 515 EPDSTGVPLIKYKAEKRVIGHEKMTKYYDAKEVNAESIITVMGLKPTSYVWKLAMN 574
QY 456 ---GHRPIISMNVN-----GTASE-----QDDPQVR-----LST 483
Db 575 GKGLGDSPTQOEFTTQVREPSAPKLVLGHLSDEGNSIKVDILKQDDGSPIRHYLVNRA 634
QY 484 LNVLV-TPEL-----LETGVE---CTASNDLGKNTSL--FLELVNLTTLT 523
Db 635 LNLKEMPEKRVPSNSHHVWLKALENNVDYVALENOGSKKPLLSFRTAKTTAT 694
QY 524 PDSNTTGLSTSTASPTTRANSTSTERKLPESRGVIVAVICILVLAIVGAVLYFLY 583
Db 695 ATASAGTGLCTG-----RRSGKQELTPPSKSELVVEYKSKLPREMKLLG 626
QY 584 KKGLKPC-----RRSGKQELTPPSKSELVVEYKSKLPREMKLLG 626
Db 732 KGLGLMCIAVNPCKGKAPGAKKGDIEBKRAFSKDESK--EPIVEVTE----- 778
QY 627 SSQDKRAPDQGEKRYD 643
Db 779 ---EERTPNHDSNQIE 792

```

```

RESULT 11
NCA2_MOUSE
ID NCA2_MOUSE STANDARD: PRT; 725 AA.
AC P13594; 061950;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE, PHOSPHATIDYLINOSITOL-LINKED ISOFORM
DE PRECURSOR (N-CAM 120) (NCAM-120).
GN NCAM1 OR NCAM.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-87246524; Pubmed-3555563;
RA Barthele D., Santoni M.-J., Wille W., Ruppert C., Caix J.-C.,
RA Hirsch M.-R., Fontecilla-Camps J.-C., Goridis C.;
RT "Isolation and nucleotide sequence of mouse NCAM cDNA that codes for
RT a Mr 79,000 polypeptide without a membrane-spanning region."
RL EMBO J. 6:907-914(1987).
RN [2]
RP SEQUENCE OF 20-700 FROM N.A.
RC STRAIN-C57BL/6; TISSUE=BRIN;
RX MEDLINE-89251563; Pubmed-2712486;
RA Santoni M.J., Barthele D., Vopper G., Boned A., Goridis C., Wille M.;
RT "Differential exon usage involving an unusual splicing mechanism
RT generates at least eight types of NCAM cDNA in mouse brain."
RL EMBO J. 8:385-392(1989).
RN [3]
RP SEQUENCE OF 642-725 FROM N.A.
RX MEDLINE-88283628; Pubmed-3396534;
RA Barbs J.A., Chaix J.C., Steinmetz M., Goridis C.;
RT "Differential splicing and alternative polyadenylation generates
RT distinct NCAM transcripts and proteins in the mouse."
RL EMBO J. 7:625-632(1988).

```

```

RN [4]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; Pubmed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RT domain of mammalian neural cell adhesion molecules."
RL J. Biol. Chem. 261:3396-3401(1986).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00051; CAA68263.1; -
DR EMBL; X15049; CAA33148.1; ALT_SEQ.
DR EMBL; X07195; CAA30173.1; -
DR PIR; A29673; IJMSNG.
DR MGD; MGI:97281; NCAM.
DR INTERPRO; IPR001777; -
DR INTERPRO; IPR003006; -
DR PFAM; PF00041; fn3; 2.
DR PFAM; PF00047; ig; 5.
KM Cell adhesion; Glycoprotein; Repeat; Alternative splicing;
KM Immunoglobulin domain; Signal; Heparin-binding; GPI-anchor.
FT SIGNAL 1 19
FT CHAIN 1 725
FT FT 20 725
FT FT 34 103
FT FT 132 196
FT FT 228 295
FT FT 323 393
FT FT 420 487
FT FT 519 596
FT FT 625 692
FT FT 152 156
FT FT 161 165
FT FT DISULFID 41 96
FT FT DISULFID 139 189
FT FT DISULFID 235 288
FT FT DISULFID 330 386
FT FT DISULFID 427 480
FT FT CARBOHYD 222 222
FT FT CARBOHYD 316 316
FT FT CARBOHYD 348 348
FT FT CARBOHYD 424 424
FT FT CARBOHYD 450 450
FT FT CARBOHYD 479 479
FT FT CARBOHYD 261 268
FT FT CONFLICT 273 273
FT FT CONFLICT 354 355
FT FT CONFLICT 549 549
FT FT CONFLICT 572 572
FT FT CONFLICT 575 575
FT FT CONFLICT 589 594
FT FT CONFLICT 600 602
FT FT CONFLICT 657 657
FT FT SEQUENCE 725 AA; 80296 MW; C2AEBB84161C6B2F CRC64;

```

Query Match 7.3%; Score 246.5; DB 1; Length 725;
 Best Local Similarity 20.6%; Pred. No. 6.1e-09;
 Matches 141; Conservative 92; Mismatches 261; Indels 189; Gaps 33;

Db 343 PSEFW-RTQTDSPINGEDEGATSTLSPVGEDEHSITCTVTCORRKLEKTIQVEVY 401
 OY 390 -----PV-----LOLHLKRE-----400
 Db 402 SFEPDELEISGLVHGRPVVNCVNPVFPDLLEIELLKGETTLNKLFRREIGTKSL 461
 OY 401 -----AGGGRCAVA-----SYPSIGLNRTQLVNVAIGPRMMAFKE 437
 Db 462 ETKSLEMTFTPTAEDTKALVCIALKLHSSQMSSEPKOROSTQTLVYNA-----P---KE 513
 OY 438 RKVVV-----KEMVNLNLSCEASGHPRPTISWN---VNGTASRODDPQRYVLSLTVLV 488
 Db 514 PTVWSSPVVEEESPVLNCTSSDGFPPKILMSKOLKNGELQPLSON-----TTLSTMA 568
 OY 489 TPPELLTGV-ECTASNDLG 506
 Db 569 T-KMEDSGIVCEGINAEG 586

RESULT 13
 NCAL_RAT STANDARD; PRT; 858 AA.
 ID NCAL_RAT
 AC P13596;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE NEURAL CELL ADHESION MOLECULE, 140 KDA ISOFORM PRECURSOR (N-CAM 140)
 DE (NCAM-140).
 GN NCAM1 OR NCAM.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRIN;
 RX MEDLINE=88059265; PubMed=3680385;
 RA Small S.J., Shull G.E., Santoni M.-J., Akeson R.;
 RT "Identification of a cDNA clone that contains the complete coding
 RT sequence for a 140-KD rat NCAM polypeptide.";
 RL J. Cell Biol. 105:2335-2345(1987).
 RN [2]
 RP SEQUENCE OF 355-364 FROM N.A.
 RX MEDLINE=90166485; PubMed=2483093;
 RA Small S.J., Haines S.L., Akeson R.A.;
 RT "Polypeptide variation in an N-CAM extracellular immunoglobulin-like
 RT fold is developmentally regulated through alternative splicing.";
 RL Neuron 1:1007-1017(1988).
 CC -!- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
 CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
 CC NEURITES, ETC.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: THE DIFFERENT TISSUE-SPECIFIC FORMS OF
 CC N-CAM ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN-TYPE III-LIKE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL, X06564; CAA29809.1;
 CC EMBL, M32611; AAA41679.1;
 CC PIR, S00846; IJRTNC.
 CC INTERPRO: IPR001777;
 CC INTERPRO: IPR003006;
 CC PFAM, PF00041; fn3; 2.
 CC PFAM, PF00047; 19; 5.
 CC Cell adhesion; Glycoprotein; Transmembrane; Repeat;

KW Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding.
 FT SIGNAL 1 19
 FT CHAIN 20 858
 FT
 FT DOMAIN 20 721
 FT TRANSMEM 722 739
 FT DOMAIN 740 858
 FT DOMAIN 34 103
 FT DOMAIN 132 196
 FT DOMAIN 228 295
 FT DOMAIN 323 403
 FT DOMAIN 430 497
 FT DOMAIN 514 615
 FT DOMAIN 616 712
 FT DOMAIN 152 165
 FT DOMAIN 161 165
 FT DISULFID 41 96
 FT DISULFID 139 189
 FT DISULFID 235 288
 FT DISULFID 330 396
 FT DISULFID 437 490
 FT CARBOHYD 222 222
 FT CARBOHYD 316 316
 FT CARBOHYD 348 348
 FT CARBOHYD 434 434
 FT CARBOHYD 460 460
 FT CARBOHYD 489 489
 SO SEQUENCE 858 AA; 94658 MW; EALA06A4EA050F6 CMC64;

Query Match 7.3%; Score 244.5; DB 1; Length 858;
 Best Local Similarity 20.9%; Pred. No. 1e-08;
 Matches 147; Conservative 92; Mismatches 245; Indels 219; Gaps 34;

OY 33 PELVEVEVGSTALKCGLSQSGNLSHVD--WFSVHKERTLIFVRQGGQSGSEGEVQ 90
 Db 26 PQSGEISVGESEKFFLC---QVAGDAKDKDISWSPGEXLS-----PNOQ 67
 OY 91 RLSTL---GDRGATLALTVQTPODERFLC--GQKRPSGEYRIQLRVYKAPPEPNIOVNP 145
 Db 68 RLSVWVNDSSSTLITINANNIDAGIKCVYARBDGTQSEAVANNKIFQ-----KLMFN 122
 OY 146 LGIPVNSKEPEE-VATCVGRNGYPIQVITWYKNGRPLKEEKN-RVHIQSSQTVESGLYT 203
 Db 123 APPTQEFKEGDAVIVCDVVSILP-PTIYMKHKGROVILKDKVREFILSNLYLOIRG--- 178
 OY 204 LOSILKAQVLKEDKAQFCELVNRLRPSGNHKKESSEVTVVPFYPRTEKYWLEPEVGLK 263
 Db 179 -----IKRTDEGTYRCE-----GRILARGE-----INFKDQIVVNPPTVQAR 217
 OY 264 E-----GDRVEIRCLADGNPPHFSISKONPSTREAEEET-----TNDNGVLVLEP 309
 Db 218 QSVVNTALNGSVTLVLCADGPRPTMTKDGERTENDEEDDEKHIISDSSSELLTINN 277
 OY 310 ARKESHGREGCG-----LDLDTMISLSEPOELLVNTVSDVRSPPAAPERQESSLTL 363
 Db 278 VDKNDAEYVCIENKAGQDASIHLYKVPKPK--ITYVE-----QQTAMELEE--QVTL 328
 OY 364 TCRAES-----SODLEFQWLRETTGVLEGRPYLOLH-----DLKR 399
 Db 329 TCBAASDPDIPSTWRTSTRNISSEKASWTRPKQETLQDHMYVRSKHARYSITLKSIOY 388
 OY 400 EAGGGRCAVAS-----VPSIGPLNFTQLVNVAIFGPPMAFKERRVWVK 443
 Db 389 TIDGEYICIASNTIIGDSQSMYLEQYAPRLQG-----PVAIVY-----TW 428
 OY 444 ENNVNLNLSCEASGHPRPTISWNVNGTASEDDQDPQRYVLSLTVLVNP--ELLEET 495
 Db 429 EGNQVNVITEGVFAVPSATISWFRDQGL-----PSSNYSNIKIYNPSPASYLEVTDPSEN 483
 OY 496 ---GVECTASNDGKNTSILFELVNLTTLPDPSNTTTCGSTASTASPHTRANSTERKL 552
 Db 484 DFGVNYCTAVNRIGQES---LEFIIYQADTPSSP-----SIDRVEPY-----SSTAOVQF 530

OY 553 PEPESRGVIVAV-----IVCLVL-----AVLGAVL 579
 DB 531 DEPEATGCVPIIKYAEKKSGLGEAMHNSKWTDAKANEGLVTTIMGLPEPTRIYALAL 590
 OY 560 YFLYKKGKLPCCRRSGKQITLPPSKSELVYEVKSDKLPEMG 622
 DB 591 -----NGK-----GLGEISATATEKFTQPVREPSAPKLEGQMG 622
 RESULT 14
 PCBM_MOUSE
 ID PGBM_MOUSE STANDARD: PRT: 3707 AA.
 AC 005793;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
 DE PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
 GN HSPG2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MELANOMA;
 RX MEDLINE=92078153; PubMed=1744087;
 RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
 RA Yamada Y., Hassell J.R.;
 RT "The complete sequence of perlecan, a basement membrane heparan
 RT sulfate proteoglycan, reveals extensive similarity with laminin A
 RT chain, low density lipoprotein-receptor, and the neural cell adhesion
 RT molecule."
 RL J. Biol. Chem. 266:22939-22947(1991).
 RN [2]
 RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=89034110; PubMed=2972708;
 RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogeli G., Sasaki M.,
 RA Yamada Y., Hassell J.R.;
 RT "Identification of cDNA clones encoding different domains of the
 RT basement membrane heparan sulfate proteoglycan."
 RL J. Biol. Chem. 263:16379-16387(1988).
 CC -1- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
 CC MEMBRANS. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
 CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
 CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
 CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
 CC SUBSTRATE FOR CELLS.
 CC -1- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
 CC DIMERS OR STELLATE STRUCTURES.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANS.
 CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 CC AND O-LINKED OLIGOSACCHARIDES.
 CC -1- SIMILARITY: CONTAINS 10 5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 15 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 15 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M77174; AAA39911.1;
 DR EMBL: J04054; AAA39899.1;
 DR EMBL: J04055; AAA39912.1;
 HSSP: P01130; IAUJ.

DR MGD: MGT.96257; HSPG2.
 DR INTERPRO: IPR000034;
 DR INTERPRO: IPR000082;
 DR INTERPRO: IPR000561;
 DR INTERPRO: IPR001438;
 DR INTERPRO: IPR001791;
 DR INTERPRO: IPR002049;
 DR INTERPRO: IPR002172;
 DR INTERPRO: IPR003006;
 DR PFAM: PF01390; SEA; 1.
 DR PFAM: PF00047; I9; 15.
 DR PFAM: PF00052; laminin_B; 3.
 DR PFAM: PF00053; laminin_EGF; 8.
 DR PFAM: PF00054; laminin_G; 3.
 DR PFAM: PF00057; Idl_receptor; 4.
 DR PRINTS: PR00010; EGFBLD.
 DR PROSITE: PS00022; EGF_1; 8.
 DR PROSITE: PS01186; EGF_2; 5.
 DR PROSITE: PS01209; LDLRA_1; 4.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 11.
 DR PROSITE: PS50068; LDLRA_2; 4.
 KW Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
 KW Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
 KW Extracellular matrix; EGF-like domain.
 FT CHAIN 1 21
 FT 22 3707
 FT 22 193
 FT 194 403
 FT 404 504
 FT 507 1676
 FT 1677 2980
 FT 2981 3707
 FT 194 234
 FT 281 319
 FT 320 359
 FT 360 403
 FT 404 504
 FT 521 530
 FT 531 730
 FT 731 763
 FT 764 813
 FT 814 871
 FT 879 923
 FT 924 933
 FT 934 1125
 FT 1126 1158
 FT 1159 1208
 FT 1209 1265
 FT 1275 1324
 FT 1325 1334
 FT 1335 1529
 FT 1530 1562
 FT 1563 1612
 FT 1613 1670
 FT 1677 1771
 FT 1772 1865
 FT 1866 1954
 FT 1955 2049
 FT 2050 2148
 FT 2149 2244
 FT 2245 2343
 FT 2344 2436
 FT 2437 2533
 FT 2533 2619
 FT 2620 2720
 FT 2721 2809
 FT 2810 2895
 FT 2896 2980
 DOMAIN II (4 LDLRA REPEATS).
 DOMAIN IIA (1 IGG-REPEAT).
 DOMAIN III (SIMILAR TO SHORT ARM OF LAMININ A CHAIN).
 DOMAIN IV (SIMILAR TO NEURAL CELL ADHESION MOLECULE; 14 IGG REPEATS).
 DOMAIN V (C-TERMINAL G-DOMAIN OF LAMININ ALPHA CHAINS AND EGF).
 LDL-RECEPTOR CLASS A 1.
 LDL-RECEPTOR CLASS A 2.
 LDL-RECEPTOR CLASS A 3.
 LDL-RECEPTOR CLASS A 4.
 IG-LIKE C2-TYPE DOMAIN 1.
 LAMININ EGF-LIKE 1 (N-TERMINAL).
 LAMININ DOMAIN IV 1 (DOMAIN III A).
 LAMININ EGF-LIKE 1 (C-TERMINAL).
 LAMININ EGF-LIKE 2.
 LAMININ EGF-LIKE 3.
 LAMININ EGF-LIKE 4 (INCOMPLETE).
 LAMININ EGF-LIKE 5 (N-TERMINAL).
 LAMININ DOMAIN IV 2 (DOMAIN III B).
 LAMININ EGF-LIKE 5 (C-TERMINAL).
 LAMININ EGF-LIKE 6.
 LAMININ EGF-LIKE 7.
 LAMININ EGF-LIKE 8.
 LAMININ EGF-LIKE 9 (N-TERMINAL).
 LAMININ DOMAIN IV 3 (DOMAIN III C).
 LAMININ EGF-LIKE 9 (C-TERMINAL).
 LAMININ EGF-LIKE 10.
 LAMININ EGF-LIKE 11.
 IG-LIKE C2-TYPE DOMAIN 2.
 IG-LIKE C2-TYPE DOMAIN 3.
 IG-LIKE C2-TYPE DOMAIN 4.
 IG-LIKE C2-TYPE DOMAIN 5.
 IG-LIKE C2-TYPE DOMAIN 6.
 IG-LIKE C2-TYPE DOMAIN 7.
 IG-LIKE C2-TYPE DOMAIN 8.
 IG-LIKE C2-TYPE DOMAIN 9.
 IG-LIKE C2-TYPE DOMAIN 10.
 IG-LIKE C2-TYPE DOMAIN 11.
 IG-LIKE C2-TYPE DOMAIN 12.
 IG-LIKE C2-TYPE DOMAIN 13.
 IG-LIKE C2-TYPE DOMAIN 14.
 IG-LIKE C2-TYPE DOMAIN 15.

```

FT DOMAIN 2961 3130 LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
FT DOMAIN 3049 3241 EGF-LIKE 1.
FT DOMAIN 3304 3495 LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
FT SITE 3558 3705 HEPARAN SULFATE (POTENTIAL).
FT SITE 65 67 HEPARAN SULFATE (POTENTIAL).
FT SITE 71 73 HEPARAN SULFATE (POTENTIAL).
FT SITE 76 78 MEDIANES MOTOR NEURON ATTACHMENT
FT SITE 3615 3617 (POTENTIAL).
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 206 225 BY SIMILARITY.
FT DISULFID 219 234 BY SIMILARITY.
FT DISULFID 285 297 BY SIMILARITY.
FT DISULFID 292 310 BY SIMILARITY.
FT DISULFID 304 319 BY SIMILARITY.
FT DISULFID 325 337 BY SIMILARITY.
FT DISULFID 332 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 368 381 BY SIMILARITY.
FT DISULFID 375 394 BY SIMILARITY.
FT DISULFID 388 403 BY SIMILARITY.
FT DISULFID 428 479 BY SIMILARITY.
FT DISULFID 764 773 BY SIMILARITY.
FT DISULFID 766 780 BY SIMILARITY.
FT DISULFID 783 792 BY SIMILARITY.
FT DISULFID 795 811 BY SIMILARITY.
FT DISULFID 814 829 BY SIMILARITY.
FT DISULFID 816 839 BY SIMILARITY.
FT DISULFID 842 851 BY SIMILARITY.
FT DISULFID 854 869 BY SIMILARITY.
FT DISULFID 1159 1168 BY SIMILARITY.
FT DISULFID 1161 1175 BY SIMILARITY.
FT DISULFID 1178 1187 BY SIMILARITY.
FT DISULFID 1190 1206 BY SIMILARITY.
FT DISULFID 1209 1224 BY SIMILARITY.
FT DISULFID 1211 1234 BY SIMILARITY.
FT DISULFID 1237 1246 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1275 1287 BY SIMILARITY.
FT DISULFID 1277 1293 BY SIMILARITY.
FT DISULFID 1295 1304 BY SIMILARITY.
FT DISULFID 1307 1322 BY SIMILARITY.
FT DISULFID 1363 1372 BY SIMILARITY.
FT DISULFID 1565 1579 BY SIMILARITY.
FT DISULFID 1582 1591 BY SIMILARITY.
FT DISULFID 1594 1610 BY SIMILARITY.
FT DISULFID 1613 1628 BY SIMILARITY.
FT DISULFID 1615 1638 BY SIMILARITY.
FT DISULFID 1641 1650 BY SIMILARITY.
FT DISULFID 1653 1668 BY SIMILARITY.
FT DISULFID 1792 1839 BY SIMILARITY.
FT DISULFID 1886 1932 BY SIMILARITY.
FT DISULFID 1976 2021 BY SIMILARITY.
FT DISULFID 2073 2118 BY SIMILARITY.
FT DISULFID 2170 2215 BY SIMILARITY.
FT DISULFID 2268 2313 BY SIMILARITY.
FT DISULFID 2365 2413 BY SIMILARITY.
FT DISULFID 2456 2506 BY SIMILARITY.
FT DISULFID 2534 2599 BY SIMILARITY.
FT DISULFID 2641 2686 BY SIMILARITY.
FT DISULFID 2831 2876 BY SIMILARITY.
FT DISULFID 2917 2962 BY SIMILARITY.
FT CARBOHYD 65 65 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT CARBOHYD 71 71 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT CARBOHYD 76 76 O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match Best Local Similarity 22.3%; Score 244.5; DB 1; Length 3707;

Matches 113; Conservative 79; Mismatches 190; Indels 125; Gaps 26;

QY 17 CCRPVAGVP-----GEAEQAPAELEVEV-----GSTALCKGLSOGNLSHVDF 63

```

Db 1653 CAPGEGDPNAGGRCODPLTKESLEVOIHSRSVYPOGGHSLRC---QVSGSPHYFW 1709
QY 64 SVHKERKTLIFRVQOGOGSEGEYEOPLSLQDGATLALTOYTPODERFLFCGR-PR 122
Db 1710 S-----REDGRPLPSSAQOR-----HGSLSHPSPVSPASGYITCTCRLLH 1752
QY 123 SQEIRQLRYKADPEEPNIOVPLGIVNSKEPEVA-----TCVRNGYPIPOVIW 175
Db 1753 TSNRAELVLAEAPSK-----PIWVTEEQRSQSVAPGADVFTICTAKSKPAYLLWMT 1806
QY 176 K--NGRPLKEKKNVHIIQSSQTVESGLTLOSTIKRLQYKENDQAFY-CEIN--YRLP 230
Db 1807 RLHNGK-----LPSADFNGLITIRV-----QPSDAGTYVCTSSNMFMAD 1848
QY 231 SGN---HMKSEREVTEV--FYPEKYLEVEPEVGMLEKEDRYEICLDGNGPPHF--- 282
Db 1849 OGATLHVQVSGSTAVASIHPPQ--LVQP-----CQQAFFKCSATGNTPTMLEMT 1899
QY 283 -SISKQNPSTREAEETNDNGVLEVPARKHSGRYECQGLD-----LDTMISLL--S 333
Db 1900 GGPSGQLPAAQIHL-----NGILRLPALEPSPDQGYLCRLSSAGHVARAMLQYHGG 1953
QY 334 EPQELVNVYSDVAVSPAPEREQSSLTLCPEASSODLEFQWLEETGVYER---G 389
Db 1954 GPR-----VOVSPEKQVHEGRTVRLCYCAAGVPSASITW--RKEGSLPFHQAHG 2003
QY 390 PVLQHLDKREKAGGYRCVAVSPISPLGNTOLVNV-----AIFGPMMAFKERKVV 441
Db 2004 SRLRLHMSYADSGEYVCGRAN--NNIDAQETSIMISVSPNSPPAPASAPRIRESSSR 2062
QY 442 VKENWVNLSCASGHPRTISMVNG 468
Db 2063 VAEQGTLDLNCVVGHAHVWTKRKG 2089

RESULT 15
NCAL BOVIN STANDARD; PRT; 853 AA.
AC P31836;
DT 01-JUL-1993 (rel. 26, Created)
DT 01-JUL-1993 (rel. 26, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE NEURAL CELL ADHESION MOLECULE, 140 KDA ISOFORM PRECURSOR (N-CAM 140)
DE (NCAM-140).
GN NCAM1 OR NCAM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE-BRAIN CORTEX.
RX MEDLINE=89378239; PubMed=2776887;
RA Lipkin V.M., Khramtsov N.V., Andreeva S.G., Moshnyakov M.V.,
RA Petukhova G.V., Rakiлина T.V., Feshchenko E.A., Ishchenko K.A.,
RA Mirzoeva S.F., Chernova M.N., Dranyatsyna S.M.;
RT "Calmodulin-independent bovine brain adenylate cyclase. Amino acid
RL sequence and nucleotide sequence of the corresponding cDNA.";
RL FEBS Lett. 251:69-73(1989).
RN [2]
RP SEQUENCE OF 20-36.
RX MEDLINE=86140120; PubMed=3512556;
RA Rougon G., Marshak D.R.;
RT "Structural and immunological characterization of the amino-terminal
RL domain of mammalian neural cell adhesion molecules.";
RL J. Biol. Chem. 261:3396-3401(1986).
RN [3]
RP IDENTIFICATION AS N-CAM.
RX MEDLINE=92111748; PubMed=1765159;
RA Premont R.T.;
RT "A bovine brain CDNA purported to encode calmodulin-insensitive
RT adenylate cyclase has extensive identity with neural cell adhesion
RT molecules (N-CAMs).";

```


Qy	1	MGRLPYCAFLLAACCCCCPRVAGVPGGEAPAPELVEYEVSTLLKCGLSQSGNLSHV	60
Qy	1	MGRLPYCAFLLAACCCCCPRVAGVPGGEAPAPELVEYEVSTLLKCGLSQSGNLSHV	60
Db	1	MGRLPYCAFLLAACCCCCPRVAGVPGGEAPAPELVEYEVSTLLKCGLSQSGNLSHV	60
Qy	61	DMFSYHKKEKPTLIPFRVGGGSGSSEPGKEEQLSLDDEGATLALQVYPPDERITFLCGGR	120
Qy	61	DMFSYHKKEKPTLIPFRVGGGSGSSEPGKEEQLSLDDEGATLALQVYPPDERITFLCGGR	120
Db	61	DMFSYHKKEKPTLIPFRVGGGSGSSEPGKEEQLSLDDEGATLALQVYPPDERITFLCGGR	120
Qy	121	PRSGEYRQLRVYKAPKEEPTNQVPLGIPVNSKDEEATCVGNGSPIPQVIMYKNGRP	180
Qy	121	PRSGEYRQLRVYKAPKEEPTNQVPLGIPVNSKDEEATCVGNGSPIPQVIMYKNGRP	180
Db	121	PRSGEYRQLRVYKAPKEEPTNQVPLGIPVNSKDEEATCVGNGSPIPQVIMYKNGRP	180
Qy	181	LKEEKNRHHIQQSQTVESSGLYTLQSLIKALVVEDKDAQCYCECLNRLDSGNMKRSRE	240
Qy	181	LKEEKNRHHIQQSQTVESSGLYTLQSLIKALVVEDKDAQCYCECLNRLDSGNMKRSRE	240
Db	181	LKEEKNRHHIQQSQTVESSGLYTLQSLIKALVVEDKDAQCYCECLNRLDSGNMKRSRE	240


```

Oy 241 VTPVPEVPTKEMVLEVEPVGMLEKEDRVEIRCLADGNPPHFSISKONPSTREAEETTN 300
    |||||
Db 241 VTPVPEVPTKEMVLEVEPVGMLEKEDRVEIRCLADGNPPHFSISKONPSTREAEETTN 300
Oy 301 DNGVLVLEPARKHSGRYECOGLDLDTMIISLSEPOELLVNVYSDRVSPAPEROEGSS 360
    |||||
Db 301 DNGVLVLEPARKHSGRYECOGLDLDTMIISLSEPOELLVNVYSDRVSPAPEROEGSS 360
Oy 361 LTTTGEAASSQDLFEQWMLREFTQVLERGPVLOHDLKREAGGGRCAVASPSTPGLNRT 420
    |||||
Db 361 LTTTGEAASSQDLFEQWMLREFTQVLERGPVLOHDLKREAGGGRCAVASPSTPGLNRT 420
Oy 421 QLVNVAIFGPPMAFKERKVVKNMVLNLSCEASGHPRTISMVNVNSTASEQODDOPRV 480
    |||||
Db 421 QLVNVAIFGPPMAFKERKVVKNMVLNLSCEASGHPRTISMVNVNSTASEQODDOPRV 480
Oy 481 LSTLVNLTPELLETGVECTASNDLGKNTSILFELVNLTTLPDSNTTGLSTASPH 540
    |||||
Db 481 LSTLVNLTPELLETGVECTASNDLGKNTSILFELVNLTTLPDSNTTGLSTASPH 540
Oy 541 TRANSTSTERKLPESRGVTVAVICILVAVLGAVLYLKKGKLPCHRSKQETTL 600
    |||||
Db 541 TRANSTSTERKLPESRGVTVAVICILVAVLGAVLYLKKGKLPCHRSKQETTL 600
Oy 601 PPSRKSELVEVSKDLPEEMGLLOGSSGDRAPDQGEKTYDLRH 646
    |||||
Db 601 PPSRKSELVEVSKDLPEEMGLLOGSSGDRAPDQGEKTYDLRH 646

```

```

RESULT 2
O98922 PRELIMINARY: PRT: 626 AA.
ID 098922:
AC 098922:
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE HEMCAM PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H-B190V+; TISSUE=THYMUS;
RX MEDLINE=9713433; PubMed=8978830;
RA Valerio O., Dunon D., Aissel F., Dangy J.P., McNaghy K.M., Imhof B.A.;
RT "HEMCAM, an adhesion molecule expressed by c-kit+ hemopoietic
RT progenitors.";
RL J. Cell Biol. 135:1655-1668(1996).
DR EMBL: Y08855; CA70080.1;
DR INTERPRO: IPR003006;
DR PFAM: PF00047; 19; 5.
KW Signal.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 626 POTENTIAL.
SQ SEQUENCE 626 AA: 69104 MW: 27882F47044E163 CRC64;

```

Query Match 32.8%; Score 1104; DB 13; Length 626;
 Best Local Similarity 38.3%; Pred. No. 6.3e-78;
 Matches 251; Conservative 116; Mismatches 221; Indels 68; Gaps 16;

```

Oy 8 CAFLLAACCCPRVAGVGEAEPAPLVEVEVSTALLKGLS-QSGNLSHVDWFSVH 66
    |||||
Db 22 CPEFLCLLCC-----GAAGRLVYMPAVLVEIGSTARLESFSTPGNASTSIEMFVFN 76
Oy 67 KEKRTLRIFRVROGQ-----OSEPGEYEQRLSLQDRGATLALTQVTPDDE-RIFLCQ 117
    |||||
Db 77 RRP-----LRSGEAVRHHSAGVRIIDETFEYSERLSVGEDKA-LTSKTRQDNARTFICQ 129
Oy 118 --GKRPRSOEYRIQLRVYKABEENPIQVNLGIPVNSKEPEEVATCVGRNGYPIPIQVIMY 175

```

```

Db 130 VGADSGVGSFRELTYTKIPAPPEITPNASGIPAGSNMLKIAQTSBNSPSPNITMY 189
    |||||
Oy 176 KNGRPLKEENNRHIOSSQVVESSGLYTLQSLKADLVEDDAQCYCELNRLPSGNIM 235
    |||||
Db 190 KNEPPLQEDDKTKILTTLVRESNGLYTVVSTLFSKVTREDRSLPHCVHWLQGMRT 249
Oy 236 KESREVTVPVPEVPTKEMVLEVEPVGMLEKEDRVEIRCLADGNPPHFSISKONPSTR 292
    |||||
Db 250 KDSPRVAVTVPTPEVVELRVATNAGIVKGGDYKLVCAADGNAPAVFSPFRELDSNQ 309
Oy 293 EAEETTNQGVLEPARKHSGRYECOGLDLDTMIISLSEPOELLVNVYSDRVV -SP 350
    |||||
Db 310 DMTSLADTNDGVLMHLNHSKSSGLYRCQTLDDDTQHEGD -VELVNVYIGVQYKMP 368
Oy 351 AAPEROGSSLTLTCEASSQDLFEQWMLREFTQVLERGPVLOHDLKREAGGGRCAVAS 410
    |||||
Db 369 SSP-LHEGDSVRLSTGASHSPVKLDYQW-RDARCRKAENQULLNLNLTETSSNSCRVK 426
Oy 411 VPSIPGLNRTQLVNVAIFGPPMAFKERKVVKNMVLNLSCEASGHPRTISMVNVNGTA 470
    |||||
Db 427 ARSVPLGEQSKQVAVAVKGRPRIVALSAPLYRQDEVINLTCKALAFQPSFHMISNGTT 486
Oy 471 SEDQDPQRYLSTLVNLTPELLETGVECTASNDLGKNTSILFELVNLTTLPDSNTT 530
    |||||
Db 487 HEY-MENQHMASNLTVRVSHDLRAGAMCRVSNALGVSE-----KHQILDQKPS----- 535
Oy 531 GLSTSTASPTTRANSTSTERKLPESRGVTVAVICILVAVLGAVLYLKKGKLP 590
    |||||
Db 536 -----ESKGIITVAITVILVAVLGAVLYLKKGKLP 570
Oy 591 RRSQKQETLPPSRKSELVEVSKDLPEEMGLLOGSSGDRAPDQGEKTYDLRH 646
    |||||
Db 571 GRSGKQDITKPEARKDKNVVEVSKSLSEAGLLOGANCKRSPAPQSKKYDILRH 626

```

```

RESULT 3
O90880 PRELIMINARY: PRT: 626 AA.
ID 090880:
AC 090880:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
DE L-GICERIN PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEGHORN; TISSUE=GIZZARD;
RX MEDLINE=96081930; PubMed=7493386;
RA Taira E., Nagino T., Tanizura H., Takahata N., Kim C., Kuo C., Li B.,
RA Higuuchi H., Miki N.;
RT "Expression and functional analysis of a novel isoform of glicerin, an
RT immunoglobulin superfamily cell adhesion molecule.";
RL J. Biol. Chem. 270:26681-26687(1995).
DR EMBL: D49849; BAA08648.1;
DR INTERPRO: IPR003006;
DR PFAM: PF00047; 19; 5.
SQ SEQUENCE 626 AA: 69074 MW: C4791EBC2EC55133 CRC64;

```

Query Match 30.5%; Score 1025; DB 13; Length 626;
 Best Local Similarity 37.4%; Pred. No. 9.4e-72;
 Matches 248; Conservative 114; Mismatches 219; Indels 82; Gaps 19;

```

Oy 8 CAFLLAACCCPRVAGVGEAEPAPLVEVEVSTALLKGLS-QSGNLSHVDW 62
    |||||
Db 22 CPEFLCLLCC-----GAAGR-----EVYMPSSA-----GSGDHEHQAQVQLLH-FW 63
Oy 63 FSVHKEKRTLIR--VROGQ-----OSEPGEYEQRLSLQDRGATLALTQVTPDDE 111

```

Db 64 ECLHLRVLVLCQPRRLSRGSEAVRHHSAGVRIDETFEYSERLSVGEDEKA-LSISKVTRQDN 122
 QY 112 -RIFLCQ--GKRRSOEYIOLRYKAPDEEPINOVNPGIPVNSKEPEEVATCVGRNGCP 168
 Db 123 ARTFICQVADSGVGESESTELTYTKIPAPPEITPNSAGIPQASNDMKIAOCTSENSEPR 182
 QY 169 IPQVITVKGGRPLKEKNVHIQSSQTVESGTYTLOSILKAOLVKEDEDAQFYCELYN 228
 Db 183 SPRTITVKGGEPLQEDKTKILITLVRESNGLYTVSTLFSKVREDRNSLPHCTVHYW 242
 QY 229 LPSGNHAKESREYTVVFPYTERKVMLEVER-VGMLEKEDREIICLADGNPPPHFSISKO 287
 Db 243 LQGMRTKQSPRVNVTVPTEHVELRVATFNAGIYKEDGVKLYCADGNAPVPSFRR 302
 QY 288 --NPSTREAEETTDNDGVVLEPARKESGRYECQGLDPTMISLSEPOELVNYVSD 345
 Db 303 ELGDSWQDMTSLADTNDGYVLMHNSKSSGLYRCOTLIDMTQHEGD-VELVYNYIEG 361
 QY 346 VRY--SPAAPERQESSLTLTCEABSSQDLFEQWLEERTGOVLERGPVQLHDLKREAG 403
 Db 362 VOYKMESSP-LHEDSVRLSTASHSPVKLDYQW-RDANGKRYABESNOLLNLTFETSS 419
 QY 404 GYRCVASVSPIGPLNLTOLVNVAFISPPMAFERKRVYKENVNLSCSEASGHPRTIS 463
 Db 420 NFSCRKARVSPVLEQSKOVAVAKPRIVASAPLYRQDEVILNLCALAPQPSFH 479
 QY 464 WNVNGTASQDQDPRVLTSLNVLTPELLETGVECTASNDLCKNTSILFELVNLTLT 523
 Db 480 WSNVGTTHHEV-MENOHMASNLTVRVSHDLRAGAMCRVSNALGVSE----KHQILDOK 533
 QY 524 PDNNTTGLSTASPHTRANSTSTERKILPEPESRGVIVAVYICVLAVGAVLYFLY 583
 Db 534 PS-----ESKGIITVAITVCLVAVAGSITVFLH 563
 QY 584 KKGLPCRRSGKOEITLPSRKSELVVEKSDKLPEEMGLQSSGDKRAPDQGEKYID 643
 Db 564 KKKGISGRSGKODITKPRKARKNVVEYKSDKLSSEAGILQCANAKESKSPADQSKYID 623
 QY 644 LRH 646
 Db 624 LRN 626

RESULT 4
 Q98921 PRELIMINARY; PRT; 584 AA.
 AC 098921;
 DT 01-FEB-1997 (TREMblrel. 02, Created)
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE HEMCAM PRECURSOR.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H.B19OV+; TISSUE=THYMUS;
 RX MEDLINE=97133433; PubMed=8978830;
 RA Valino O., Duncan D., Aisai F., Dancy J.P., McNagay K.M., Imhof B.A.;
 RT "HEMcam, an adhesion molecule expressed by c-kit+ hemopoietic
 progenitors";
 RL J. Cell Biol. 135:1655-1668(1996).
 DR EMBL: Y08856; CAA70081.1;
 DR INTERPRO: IPR003006;
 DR PRAM: PF00047; 1g; 5.
 KW Signal.
 FT SIGNAL 1 35 POTENTIAL.
 FT CHAIN 36 584 POTENTIAL.
 RN SEQUENCE 584 AA; 64422 MW; 5972D946604AF6F3 CRC64;

Query Match 27.8%; Score 936; DB 13; Length 584;
 Best Local Similarity 35.9%; Pred. No. 7, 7e-65;
 Matches 218; Conservative 109; Mismatches 213; Indels 68; Gaps 16;

QY 8 CAPLLAACCCPRVAGVGEADAPAPLVEVEVSTALCKGIS-OSQNLSHVDFSVH 66
 Db 22 CPFLCLLCC-----GAAGRLEVMPVAVLEVEISTARLECSFIPQNASFTSIEMFYV 76
 QY 67 KEKRLIFRVQSG-----OSEPGEYQRLSDRGTALATVYQDE-RIFLCQ 117
 Db 77 RRP-----LKSGEAVRHHSAGVRIDETFEYSERLSVGEDEKA-LSISKVQDNARFTICQ 129
 QY 118 --GKRRSOEYIOLRYKAPDEEPINOVNPGIPVNSKEPEEVATCVGRNGCP 175
 Db 130 VGADSGVGESESTELTYTKIPAPPEITPNSAGIPQASNDMKIAOCTSENSEPSFRTW 189
 QY 176 KNGRPLKEKNRVHIQSSQTVESGTYTLOSILKAOLVKEDEDAQFYCELYNRLPSGNH 235
 Db 190 KNGEPLQEDKTKILITLVRESNGLYTVSTLFSKVREDRNSLPHCTVHYWLOGMRT 249
 QY 236 KESREYTVPVYTERKVMLEVER-VGMLEKEDREIICLADGNPPPHFSISKO--NPSTR 292
 Db 250 KDSRVNVTVPTEHVELRVATFNAGIYKEDGVKLYCADGNAPVPSFRRRLGDSWQ 309
 QY 293 PAEETTDNDGVVLEPARKESGRYECQGLDPTMISLSEPOELVNYVSDVRY--SP 350
 Db 310 DMTSLADTNDGYVLMHNSKSSGLYRCOTLIDMTQHEGD-VELVYNYIEGQVMEP 368
 QY 351 AAPEROGSSSLTLTCEABSSQDLFEQWLEERTGOVLERGPVQLHDLKREAGGYRCVAS 410
 Db 369 SSP-LHEDSVRLSTASHSPVKLDYQW-RDANGKRYABESNOLLNLTFETSSNFSCRK 426
 QY 411 VPSIPGLNLTOLVNVAFISPPMAFERKRVYKENVNLSCSEASGHPRTISNNVNGTA 470
 Db 427 AASVGLQDSQKOVAVAKPRIVASAPLYRQDEVILNLCALAPQPSFHHISNGTT 486
 QY 471 SQDQDQDPRVLTSLNVLTPELLETGVECTASNDLCKNTSILFELVNLTLTTPDSMTT 530
 Db 487 HEY-MENOHMASNLTVRVSHDLRAGAMCRVSNALGVSE-----KHQILDOKS----- 535
 QY 531 GLSTASPHTRANSTSTERKILPEPESRGVIVAVYICVLAVGAVLYFLYKRLPC 590
 Db 536 -----ESKGIITVAITVCLVAVAGSITVFLHKKKISG 570
 QY 591 RRSKOEI 598
 Db 571 GRSKODI 578

RESULT 5
 Q90989 PRELIMINARY; PRT; 584 AA.
 AC 090989;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
 DE S-GICERIN PRECURSOR.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LEGHORN; TISSUE=GIZZARD;
 RX MEDLINE=94213753; PubMed=8161457;
 RA Taira E., Takaha N., Taniguchi H., Kim C., Miki N.;
 RT "Molecular cloning and functional expression of gicerin, a novel cell
 adhesion molecule that binds to neurite outgrowth factor";
 RL Neuron 12:861-872(1994).
 RN [2]
 RN SEQUENCE FROM N.A.

RC STRAIN=LEGHORN; TISSUE=GIZZARD;
RA MEDLINE=96081930; PubMed=7493388;
RA Taira E., Nagino T., Tanlura H., Takaha N., Kim C., Kuo C., Li B.,
RA Higuchi H., Miki N.;
RT "Expression and functional analysis of a novel isoform of gicerin, an
RT immunoglobulin superfamily cell adhesion molecule.";
RT J. Biol. Chem. 270:28681-28687(1995).
DR EMBL: D38559; BAA07563.1; -;
DR INTERPRO: IPR003006; -;
DR PFAM: PF00047; 19; 5;
KM Cell adhesion: Extracellular matrix.
SQ SEQUENCE 584 AA; 64378 MW; 876E0C3E920BA92F CRC64;

Query Match 25.7%; Score 863; DB 13; Length 584;
Best Local Similarity 35.1%; Pred. No. 3.9e-59;
Matches 216; Conservative 107; Mismatches 210; Indels 82; Gaps 19;

QY 8 CAFLLAACCCPRVAGVGEAPAPALVEVEVSTALCKGL-----SOSGMLSHVDV 62
DB 22 CPELLCLCC-----GAAGRP-----EYMPSSA-----GSGDREHSGVQLLH-PW 63
QY 63 FSVHKEKRTLLFR--VRGOG-----QSEPEYEQRLSLDRGATLALQVTPDE 111
DB 64 ECLHLRNVLCQPRPLRSGEAVRHNASGVRIDETYSERLSVGEDKA-LSISKVTRQDN 122
QY 112 -RIPLCQ--GKRPSOEYRIQLRVYKAPKEPNIOVNLGIPVNSKEPEEVAATCVGRNGYP 168
DB 123 ARTFCIQVAGDSOGVGESERTLYTKIPAPPEITPNSAGIPQSDNMLKIAQCTSENSEFP 182
QY 169 IPOVITWNGRPLKEEKRVHIOSSQTVESGLTLOSILKAQLVKEDKQAFCELNRY 228
DB 183 SPNTITWNGEPLQDEEKTILTTLVRESGLYTVSTLFSEKTRDRNSLFHCTYHW 242
QY 229 LPSGNHMKSEVTVYVPEPTKEVYLEVER-VGMLKEDGRVEICLADGNPPHFSISKQ 287
DB 243 LOGQWRTKDSRVNVTVYPTHEVHELVAITAGIVKEDDVKLVCDDADGNAPAFSEFFRR 302
QY 288 --NSTRAEETTNDNGVLVLEPAKKEHSGRYCOGLDITMISLSEPOELLVNVSD 345
DB 303 ELGDSWDMTSLADTNDVLMHLNVSKSSGLYRCQTLDDMTQHEGD-VELVNVYIEG 361
QY 346 VRV--SPAAREGSSLTITLCEASSQDLEFQWLRRETOGVLENGRVLQHLDRKAG 403
DB 362 VOVMKESSP-LHEGDSVRLSTAHSPVKLDYQW-RDARGKVAEGNQLLTITLTFETSS 419
QY 404 GYRCVASVPSLIGLNTQLVNVAFGPPMAFKERKYVMKENVNLNLSCEASGHPRPTIS 463
DB 420 NFSCRVARSVPLGEOSKOVAVAVKGPRIVAISAPLYVQDEVINLTCKAIAPQPSFH 479
QY 464 MNVNGTASSEDQDPQVRLSTLNVLTPELLETGYECTASNDLKNNSILFLELVNLTTLT 523
DB 480 WSVNGTTHYH-MENQHNASNLTVVSHDLRAGAMCRVSNALGVSE-----KHQLDLQK 533
QY 524 PDSNTTGLSTASPTTRANSTSTERKLPEPESRGVIVAVIYCLIVLAVLAVLEYEL 583
DB 534 PS-----ESKGIITVAIIYCLIVLAVLAVLAVLAVLAVLAVLAVLAVLAVL 563
QY 584 KKGKLPGRSGKQEI 598
DB 564 KKGKISGCRSGKQDI 578

RESULT 6
Q98923 PRELIMINARY; PRT; 504 AA.

AC Q98923;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE HEMCAM PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H.B190V+; TISSUE=THYMUS;
RA MEDLINE=97133433; PubMed=8978830;
RA Valino O., Dunon D., Aissi F., Dangy J.P., McNagny K.M., Imhof B.A.;
RT "HEMCAm, an adhesion molecule expressed by c-kit+ hemopoietic
RT progenitors.";
RT J. Cell Biol. 135:1655-1668(1996).
DR EMBL: Y08854; CAA70079.1; -;
DR INTERPRO: IPR003006; -;
DR PFAM: PF00047; 19; 5;
KM Signal.
FT CHAIN 1 35 POTENTIAL.
FT SIGNAL 36 504 POTENTIAL.
SQ SEQUENCE 504 AA; 55540 MW; ADAE4B94ED4F02E1 CRC64;

Query Match 22.2%; Score 747; DB 13; Length 504;
Best Local Similarity 35.9%; Pred. No. 3.7e-50;
Matches 172; Conservative 91; Mismatches 184; Indels 32; Gaps 13;

QY 8 CAFLLAACCCPRVAGVGEAPAPALVEVEVSTALCKGLS--OSGMLSHVDVFSVH 66
DB 22 CPELLCLCC-----GAAGRLVYMPAVLVELEGSTARLECSFSIPGNMSTSIEMFYVN 76
QY 67 KEKRTILFRGOG-----QSEPEYEQRLSLDRGATLALQVTPDE-RIPLCQ 117
DB 77 RRP-----LRSGEAVRHNASGVRIDETYSERLSVGEDKA-LSISKVTRQDNARTFIQ 129
QY 118 --GKRPSOEYRIQLRVYKAPKEPNIOVNLGIPVNSKEPEEVAATCVGRNGYP 175
DB 130 VQADSGVGESERTLYTKIPAPPEITPNSAGIPQSDNMLKIAQCTSENSEFPSPNTITW 189
QY 176 KNGRPLKEEKRVHIOSSQTVESGLYTLLOSILKAQLVKEDKQAFCELNRYLPSGNHM 235
DB 190 KNGEPLQDEEKTILTTLVRESGLYTVSTLFSEKTRDRNSLFHCTYHWLQGMRT 249
QY 236 KESRETVVPEPTKEVYLEVER-VGMLKEDGRVEICLADGNPPHFSISKQ--NSTR 292
DB 250 KDSRVNVTVYPTHEVHELVAITAGIVKEDDVKLVCDDADGNAPAFSEFFRELQDSWQ 309
QY 293 EAEETTNDNGVLVLEPAKKEHSGRYCOGLDITMISLSEPOELLVNVSDVAV--SP 350
DB 310 DMTSLADTNDVLMHLNVSKSSGLYRCQTLDDMTQHEGD-VELVNVYIEGVQVKNMP 368
QY 351 AAPREGSSLTITLCEASSQDLEFQWLRRETOGVLENGRVLQHLDRKAGGGRVAVS 410
DB 369 SSP-LHEGDSVRLSTAHSPVKLDYQW-RDARGKVAEGNQLLTITLTFETSSNFSCHRVK 426
QY 411 VPSITPLNTQLVNVAFGPPMAFKERKYVMKENVNLNLSCEASGHPRPTISNVNT 469
DB 427 ANSVGLEOSKOVAVAVKGPRIVAISAPLYVQDEVINLTCKAIAPQPSFHMSINGT 485

RESULT 7
Q9JKB2 PRELIMINARY; PRT; 622 AA.

AC Q9JKB2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE LUTHERAN GLYCOPROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Lee G., Willig T.-N., Parsons S.F., Anstee D.J., Mohandas N.,

RA Chasis J.A.;
 RT "Mouse Lutheran glycoprotein Gene";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF246667; AAF61742.1; -
 SQ SEQUENCE 622 AA; 67543 MW; 3A877738D96F6CD9 CRC64;

Query Match 18.4%; Score 618; DB 11; Length 622;
 Best Local Similarity 28.5%; Pred. No. 6,1e-40;
 Matches 191; Conservative 102; Mismatches 275; Indels 102; Gaps 25;

QY 10 FLAACCPCPVAGVGEAEPAPLVEVEVSTALLKCGLSQSGNLSH--VDWF--- 63
 DB 15 FLTSG-----YSGAQLHVSVPPEVMEVGEVALLDC---TPREHEHVLEMFVLDG 65
 QY 64 -----SVHKEKRTLIFRVKOGGSEPEYEQRLSLDRGATLALTQVTPDDEITFL 115
 DB 66 TGARHRLASVPEQSEFGLTV-HSLGRVPPEVDSR-----GRLVIKAVQVGDGRDYV 117
 QY 116 CQGR--RRSQEYRIQLRVYKAPPEPNIOVNPGLIPVNSKEPEEATCGVNGYPIPOVI 173
 DB 118 CVVAGAAGTSEATSSVAVFATPEDEYSPKGTLSVMDQPAQELATCSSNNGNPNPRTIT 177
 QY 174 WYKNGRPLK--EKKNRVHIQSQTVESSGLYTLQSLKAOLVKEDKDAQFCELYNRLP 230
 DB 178 WYRNGQRLEVPMEVNOVKYITIRTVREASGLYSLTFLYLRHKDDRDADPHCAHYDLP 237
 QY 231 SGNHMK-ESREVTVPVPTPKV--WLEVEPV--GMLKEGRVETRICLADGNPPPHSIS 285
 DB 238 SGQGRDLSTHFRLTLHPTHEVFEVWGSPTTEGWRREGDAVOLLCOGDSSPSPEYSFF 297
 QY 286 KQNSTREAEETINDNGVLVLEPARKHSGRYEC--OGLDLDMISLSPEOELLVYV 343
 DB 298 KQO-GTQE-EOLNVNKLKGNLTLEGVHRNOSGYGCRVEDYADAEVOLYVK-LKLHVAYL 354
 QY 344 SDVRVSPAPR---QEGSSLTLTCEASSQDLEFQWLREBTGOVLERGPVLOLHDKR 399
 DB 355 DPLELS--APELEFVFLNSSSTVNCASARGLPTVWRMTKDSV--TLADGPMLSQSVTF 410
 QY 400 EAGGVCVAVSPSTPGINRTQLVNVAIFGPPM-----AKERKRVYKEMVNLNLSCEA 454
 DB 411 DSAGTYCEASTPVPLLSRTQSFOLIVQGAPELKPNEIMPKSGNSWTEGDEV-LTCSA 469
 QY 455 SGHPRPTISWNVNGTASEODDOPR--VLSTLVNLYVPELLETGVECTASNDLKNITSIL 512
 DB 470 RGFPEPKLWMSQRGDTAPPEPREGCKSSSLMKVTSALREGVSCASINHGKGVNF 529
 QY 513 FLELVNLTTLTPDSNTTGLSTASPHTRANSTSTERKLPPEPESRGVIVAVIYCIYVL 572
 DB 530 HFGSV-----APQTAQAGVAVMAVAVSVGLL 555
 QY 573 AVLGAVALYFLYKKKGLCKRRSGKQETILPPSRKSELVEVNSDKLPEMGLQG--SSGD 630
 DB 556 LLVVAALFYCMRRKGRPGCCR--RAEKGAAPAREPEL--SHSGSERPHTGLMGPSGGG 611
 QY 631 KRAPDQGEK 640
 DB 612 RGSSEGGFDE 621

RESULT 8
 Q9R069 PRELIMINARY; PRT; 622 AA.
 AC Q9R069;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE LUTHERAN GLYCOPROTEIN.
 GN GPIV OR LV.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Rahuel C., Colin Y., Goossens D., Gane P., El Nemer W., Carrion J.P.,
 RA Le Van Kim C.;
 RT "Characterization of a mouse laminin receptor gene homologous of the
 RT human blood group Lutheran gene";
 RL Immunogenetics 0:0-0(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Parsons S.F., Lee G., Chasis J.A., Tanner M.J.A., Anstee D.J.;
 RT "Mouse Lutheran glycoprotein";
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF109160; AAF14226.1; -
 DR EMBL: AF221507; AAF34657.1; -
 DR INTERPRO: IPR003006; -
 DR PFMW; PF00047; 19; 5.
 SQ SEQUENCE 622 AA; 67670 MW; 257903F08D47EB4A CRC64;

Query Match 18.3%; Score 614; DB 11; Length 622;
 Best Local Similarity 28.4%; Pred. No. 1,3e-39;
 Matches 190; Conservative 102; Mismatches 276; Indels 102; Gaps 25;

QY 10 FLAACCPCPVAGVGEAEPAPLVEVEVSTALLKCGLSQSGNLSH--VDWF--- 63
 DB 15 FLTSG-----YSGAQLHVSVPPEVMEVGEVALLDC---TPREHEHVLEMFVLDG 65
 QY 64 -----SVHKEKRTLIFRVKOGGSEPEYEQRLSLDRGATLALTQVTPDDEITFL 115
 DB 66 TGARHRLASVPEQSEFGLTV-HSLGRVPPEVDSR-----GRLVIKAVQVGDGRDYV 117
 QY 116 CQGR--RRSQEYRIQLRVYKAPPEPNIOVNPGLIPVNSKEPEEATCGVNGYPIPOVI 173
 DB 118 CVVAGAAGTSEATSSVAVFATPEDEYSPKGTLSVMDQPAQELATCSSNNGNPNPRTIT 177
 QY 174 WYKNGRPLK--EKKNRVHIQSQTVESSGLYTLQSLKAOLVKEDKDAQFCELYNRLP 230
 DB 178 WYRNGQRLEVPMEVNOVKYITIRTVREASGLYSLTFLYLRHKDDRDADPHCAHYDLP 237
 QY 231 SGNHMK-ESREVTVPVPTPKV--WLEVEPV--GMLKEGRVETRICLADGNPPPHSIS 285
 DB 238 SGQGRDLSTHFRLTLHPTHEVFEVWGSPTTEGWRREGDAVOLLCOGDSSPSPEYSFF 297
 QY 286 KQNSTREAEETINDNGVLVLEPARKHSGRYEC--OGLDLDMISLSPEOELLVYV 343
 DB 298 KQO-GTQE-EOLNVNKLKGNLTLEGVHRNOSGYGCRVEDYADAEVOLYVK-LKLHVAYL 354
 QY 344 SDVRVSPAPR---QEGSSLTLTCEASSQDLEFQWLREBTGOVLERGPVLOLHDKR 399
 DB 355 DPLELS--VPELEFVFLNSSSTVNCASARGLPTVWRMTKDSV--TLADGPMLSQSVTF 410
 QY 400 EAGGVCVAVSPSTPGINRTQLVNVAIFGPPM-----AKERKRVYKEMVNLNLSCEA 454
 DB 411 DSAGTYCEASTPVPLLSRTQSFOLIVQGAPELKPNEIMPKSGNSWTEGDEV-LTCSA 469
 QY 455 SGHPRPTISWNVNGTASEODDOPR--VLSTLVNLYVPELLETGVECTASNDLKNITSIL 512
 DB 470 RGFPEPKLWMSQRGDTAPPEPREGCKSSSLMKVTSALREGVSCASINHGKGVNF 529
 QY 513 FLELVNLTTLTPDSNTTGLSTASPHTRANSTSTERKLPPEPESRGVIVAVIYCIYVL 572
 DB 530 HFGSV-----APQTAQAGVAVMAVAVSVGLL 555
 QY 573 AVLGAVALYFLYKKKGLCKRRSGKQETILPPSRKSELVEVNSDKLPEMGLQG--SSGD 630
 DB 556 LLVVAALFYCMRRKGRPGCCR--RAEKGAAPAREPEL--SHSGSERPHTGLMGPSGGG 611
 QY 631 KRAPDQGEK 640
 DB 612 RGSSEGGFDE 621

RESULT 9

035112 PRELIMINARY; PRT; 583 AA.
 ID 035112: 055172;
 AC 035112: 055172;
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE CD166 ANTIGEN PRECURSOR (ACTIVATED LEUKOCYTE-CELL ADHESION MOLECULE)
 DE (ALCAM) (MEMB PROTEIN) (HB2) (KG-CAM).
 GN ALCAM.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN RN
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=97347476; PubMed=9201982;
 RA Matsumoto A., Mitchell A., Kurata H., Pyle L., Kondo K., Itakura H.,
 RA Flidge N.;
 RT "Cloning and characterization of HB2, a candidate high density
 RT lipoprotein receptor. Sequence homology with members of the
 RT immunoglobulin superfamily of membrane proteins.";
 RT J. Biol. Chem. 272:16778-16782(1997).
 RN RN
 RP SEQUENCE OF 270-361 FROM N.A.
 RC TISSUE=MESENCHYME;
 RX MEDLINE=98215351; PubMed=9556065;
 RA Bruder S.P., Ricalton N.S., Boynton R.E., Connolly T.J., Jaiswal N.,
 RA Zala J., Barry F.P.;
 RT "Mesenchymal stem cell surface antigen SB-10 corresponds to activated
 RT leukocyte cell adhesion molecule and is involved in osteogenic
 RT differentiation.";
 RT J. Bone Miner. Res. 13:655-663(1998).
 RN RN
 RP SEQUENCE OF 29-38.
 RX MEDLINE=94272944; PubMed=8004458;
 RA Peduzzi J.D., Irwin M.H., Geisert E.E. Jr.;
 RT "Distribution and characteristics of a 90 kda protein, KG-CAM, in the
 RT rat CNS.";
 RT Brain Res. 640:296-307(1994).
 CC CC
 CC -1- FUNCTION: CELL ADHESION MOLECULE THAT BINDS TO CD6. INVOLVED IN
 CC NEURITE EXTENSION BY NEURONS VIA HETEROPHILIC AND HOMOPHILIC
 CC INTERACTIONS. MAY PLAY A ROLE IN THE BINDING OF T AND B CELLS TO
 CC ACTIVATED LEUKOCYTES, AS WELL AS IN INTERACTIONS BETWEEN CELLS OF
 CC THE NERVOUS SYSTEM. MAY BE INVOLVED IN OSTEOGENESIS. MAY PLAY AN
 CC IMPORTANT ROLE IN THE DEVELOPMENT OF A VARIETY OF MESENCHYMAL
 CC TISSUES (BY SIMILARITY).
 CC CC
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: STRONGEST EXPRESSION IN THE LUNG, THEN BRAIN,
 CC LIVER, AND KIDNEY. PRESENT IN THE SOMATOSENSORY SYSTEM, BASAL
 CC GANGLIA, CORTEX, OLFACTORY SYSTEM, AND CIRCUMVENTRICULAR ORGANS.
 CC CC
 CC -1- PTM: THE N-TERMINAL IS BLOCKED.
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 3
 CC C2-LIKE AND 2 V-LIKE DOMAINS.
 CC DR EMBL; AB008538; BAA23279.1; -;
 DR EMBL; Y13241; CA73693.1; -;
 DR EMBL; Y13240; CA73692.1; -;
 DR HSSP; Q13740; IKJC.
 DR INTERPRO: IPR003006; -;
 DR PFAM; PF00047; 19; 5.
 KW Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane;
 KW Signal; Antigen.
 FT SIGNAL 1 27
 FT CHAIN 28 583
 FT DOMAIN 28 527
 FT TRANSMEM 528 548
 FT DOMAIN 549 583
 FT DOMAIN 36 120
 FT DOMAIN 150 227
 FT DOMAIN 263 320
 FT DOMAIN 347 399
 FT DOMAIN 428 492
 FT DISULFD 43 113
 FT DISULFD 157 220
 FT POTENTIAL.

FT DISULFD 270 313 POTENTIAL.
 FT DISULFD 354 392 POTENTIAL.
 FT DISULFD 435 485 POTENTIAL.
 FT CONFLICT 339 339 S -> G (IN REF. 2).
 FT CARBOHYD 95 95 POTENTIAL.
 FT CARBOHYD 167 167 POTENTIAL.
 FT CARBOHYD 265 265 POTENTIAL.
 FT CARBOHYD 306 306 POTENTIAL.
 FT CARBOHYD 361 361 POTENTIAL.
 FT CARBOHYD 457 457 POTENTIAL.
 FT CARBOHYD 480 480 POTENTIAL.
 FT CARBOHYD 499 499 POTENTIAL.
 SQ SEQUENCE 583 AA; 65021 MW; D36B73854F5DE61E CRC64;
 Query Match 14.9%; Score 499.5; DB 11; Length 583;
 Best Local Similarity 24.6%; Pred. No. 1e-30;
 Matches 153; Conservative 122; Mismatches 273; Indels 73; Gaps 25;
 QY 5 RLVCAPFLAACCCCPRVAGVGEAQPAPLVEVEGSTALKKGLSGSGNLSHVDWFS 64
 DB 10 RLVCCLLSAVALRPLGLGW-----TVNSAVGDTIVMPCRIDVQ-NLMFGKMY 58
 QY 65 VHKERTLIRVRGQGS-----EPGEYQRLSLQDRGATLITQYTPQDERIFLCQ-GK 119
 DB 59 EKPDSVFLAFRSTKSKSYQYDVPDEYKDRLSLSE-NYTLSSINNAKISDEKRFVCLVT 117
 QY 120 RPRSGEIRIDLRVKAPEENIOVNPGLIPNSKEPEVATCYGRGYPPIQVWYKNGR 179
 DB 118 EDNVFEAPLVKVFQKSKPEI-VNRAAF-LETEQLKADDCISRSYDGNITWYRNRK 175
 QY 180 PLKEKNRVHIQSSQYVE-SSGLYTLQSLKAQVLEDKDAQVCELYNRLPSGNMKES 238
 DB 176 VLQPVDEVELSKKEKEDPQOLYTMSSLEYTTSDIDMPTCGVITYGSGQTIYS 235
 QY 239 REYTVPVFYTEKYLEV-EPVGLKEGDRVELRCLADGNPPH---FSISKONPSTREA 294
 DB 236 EOALFDIYPTQEVITQVLPKNAIKEGDNITQCLGNGPNPPEEFMYLPQGAEGIRSS 295
 QY 295 EEEFTNNGVLVEPARKHSGRYECQGLDLDIMISLSPQSLVNVYSVAVPAAV- 353
 DB 296 NYTVLTD-----VARNATGDVCKSLDQRMMASTT---ITVHYL-DLSLNSGEV 342
 QY 354 EROGSSLTITCEAESSQDLFQMLRETFQVLERGPVLQDLKREAGGYRCVAVSPS 413
 DB 343 TKQIGDPLPVSCITISARNNTVWMMKNIR-LRSSP-SFSSLHQDAGNYCETALQE 398
 QY 414 IPGINTQLVNVAIFGPPMAEFKERYVAKENVNLNLSCEBAGCHPPTISMNVTAS-- 471
 DB 399 VEGIKKRESLTLLVEGKP-QIKMTKKTGPSGLSKTILICVEGFPPALQWITGSGSVI 456
 QY 472 -EDQDDQ--RYLSTNLVLPPELLETGEVCTASNDLKNSTILFLELVNLTTLTPDSN 527
 DB 457 NQTESPYINGRYS--KIITSPENVT-LTCTAENQLETVNSLVASIST--PE-- 507
 QY 528 TTGLSTSTASPHTRANSTSTERKLEPEESRGVIVAVIICILVAVLAVFLY-KRG 586
 DB 508 -----HDEADDISDENR-EKVNDQAKLIGIYVGLLALVAGVYVWLTKMS 554
 QY 587 KLPGRSGKQETLTPSRKSE 607
 DB 555 KTSKSHVKNKDLGNMEENKLE 575
 RESULT 10
 ID 046634 PRELIMINARY; PRT; 521 AA.
 AC 046634;
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE CD166 ANTIGEN (ACTIVATED LEUKOCYTE-CELL ADHESION MOLECULE) (ALCAM)
 DE (FRAGMENT).

GN ALCAM.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=MESENCHYME; PubMed=9556065;
RX MEDLINE=98215351; PubMed=9556065;
RA Bruder S.P., Ricalton N.S., Boynton R.E., Connolly T.J., Jaiswal N.,
Zaia J., Barry F.P.;
RT "Mesenchymal stem cell surface antigen SB-10 corresponds to activated
leukocyte cell adhesion molecule and is involved in osteogenic
differentiation.";
RL J. Bone Miner. Res. 13:655-663(1998).
CC -1- FUNCTION: CELL ADHESION MOLECULE THAT BINDS TO CD6. INVOLVED IN
NEURITE EXTENSION BY NEURONS VIA HETEROPHILIC AND HOMOPHILIC
INTERACTIONS. MAY PLAY A ROLE IN THE BINDING OF T AND B CELLS TO
ACTIVATED LEUKOCYTES. AS WELL AS IN INTERACTIONS BETWEEN CELLS OF
THE NERVOUS SYSTEM. MAY BE INVOLVED IN OSTEOGENESIS. MAY PLAY AN
IMPORTANT ROLE IN THE DEVELOPMENT OF A VARIETY OF MESENCHYMAL
TISSUES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 3
C2-LIKE AND 2 V-LIKE DOMAINS.
DR EMBL: Y13242; CAAT3694.1; -.
DR HSSP: Q13740; IKJC.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00047; 19; 4.
KW Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane.
FT NON_TER 1 466 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 467 487 POTENTIAL.
FT DOMAIN 488 521 CYTOPLASMIC (POTENTIAL).
FT DOMAIN <1 58 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 88 165 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 201 259 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 285 337 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 366 430 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID <1 51 POTENTIAL.
FT DISULFID 95 158 POTENTIAL.
FT DISULFID 208 251 POTENTIAL.
FT DISULFID 292 330 POTENTIAL.
FT DISULFID 373 423 POTENTIAL.
FT CARBOHYD 33 33 POTENTIAL.
FT CARBOHYD 105 105 POTENTIAL.
FT CARBOHYD 244 244 POTENTIAL.
FT CARBOHYD 299 299 POTENTIAL.
FT CARBOHYD 395 395 POTENTIAL.
FT CARBOHYD 418 418 POTENTIAL.
FT CARBOHYD 437 437 POTENTIAL.
SQ SEQUENCE 521 AA; 38106 MW; A3616E9A2429E7B6 CRC64;

Query Match 14.4%; Score 485; DB 6; Length 521;
Best Local Similarity 26.4%; Pred. No. 1,2e-29;
Matches 143; Conservative 107; Mismatches 220; Indels 72; Gaps 22;

QY 314 HSGRYECGLDLDIMISLSEPELVNYSVAVSPAP--EROGSSLTLLTCEASSOD 372
DB 245 AMGDYCKSLIDKSKMSIASTA---ITVHYL-DLSLBSGEVTKOIGALLPVSCITASRN 299
QY 373 LEFOWLRETEGVLERGVVLQHLDKREAGGGRVCAVSPISPLNLTOLVNAIFGPPW 432
DB 300 ATVVMAMKDNIR--LRSSP--SFSSIQYODAGNYCETALQVEBLLKRESLTLLVEKP- 354
QY 433 MAFKERKVVYVENWVLNLSCEASGHPPTISMNNGTAS---EDQDPQ---RYLSLNV 486
DB 355 -QIKMTKKTDPGSLKTLTICHEGEPKPAIOWTITGSGSVINOTEEESPYNIGRYYST--I 411
QY 487 LVTPELEGVGCTASNDLKNSTLTFLELVNLTTLPPDSNTTGLSTASPTPRNST 546
DB 412 INSPEENVT-LTGVENDLERTVNSALSI-----PE-----HDEADEI 452
QY 547 STERKLPEPESRGVIVAVICILVAVGAVLYLY-KKGLPCRRSGKQETLTPSRK 605
DB 453 SDENR-EQVNHKATLIVGLVRLHGLAVGVWLYYKSKTASKHYNKDGLNEKK 511
QY 606 SE 607
DB 512 LE 513

RESULT 11
046651 PRELIMINARY; PRT; 521 AA.
ID 046651;
AC 01-JUN-1998 (TREMBlrel. 06, created)
DT 01-JUN-1998 (TREMBlrel. 06, last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, last annotation update)
DE CD166 ANTIGEN (ACTIVATED LEUKOCYTE-CELL ADHESION MOLECULE) (ALCAM)
DE (FRAGMENT).
GN ALCAM.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=MESENCHYME; PubMed=9556065;
RX MEDLINE=98215351; PubMed=9556065;
RA Bruder S.P., Ricalton N.S., Boynton R.E., Connolly T.J., Jaiswal N.,
Zaia J., Barry F.P.;
RT "Mesenchymal stem cell surface antigen SB-10 corresponds to activated
leukocyte cell adhesion molecule and is involved in osteogenic
differentiation.";
RL J. Bone Miner. Res. 13:655-663(1998).
CC -1- FUNCTION: CELL ADHESION MOLECULE THAT BINDS TO CD6. INVOLVED IN
NEURITE EXTENSION BY NEURONS VIA HETEROPHILIC AND HOMOPHILIC
INTERACTIONS. MAY PLAY A ROLE IN THE BINDING OF T AND B CELLS TO
ACTIVATED LEUKOCYTES. AS WELL AS IN INTERACTIONS BETWEEN CELLS OF
THE NERVOUS SYSTEM. MAY BE INVOLVED IN OSTEOGENESIS. MAY PLAY AN
IMPORTANT ROLE IN THE DEVELOPMENT OF A VARIETY OF MESENCHYMAL
TISSUES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS 3
C2-LIKE AND 2 V-LIKE DOMAINS.
DR EMBL: Y13243; CAAT3695.1; -.
DR HSSP: Q13740; IKJC.
DR INTERPRO: IPR003006; -.
DR PFAM: PF00047; 19; 4.
KW Cell adhesion; Immunoglobulin domain; Glycoprotein; Transmembrane.
FT NON_TER 1 466 EXTRACELLULAR (POTENTIAL).
FT TRANSLEM 467 487 POTENTIAL.
FT DOMAIN 488 521 CYTOPLASMIC (POTENTIAL).
FT DOMAIN <1 58 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 88 165 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 201 259 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 285 337 IG-LIKE C2-TYPE DOMAIN.

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RA Alais S., Allouli N., Vigneron P., Dunon D.;
RT "HEMCAm is the avian homologue of MCAM/CD146.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ297451; CAB97363.1; -.
FT NON-TER
SQ SEQUENCE 88 AA; 9588 MW; AAF4ADB860E0D3B CRC64;

Query Match 12.1%; Score 407; DB 11; Length 88;
Best Local Similarity 90.9%; Pred. No. 1,1e-24;
Matches 80; Conservative 2; Mismatches 6; Indels 0; Gaps 0.

QY 559 GVVAVAVICILVLAVGAVLYFLYKKGKLPGRSGKEITLPPSRKSELVEVKSCLKP 618
DB 1 GVVAVAVICTVLAVGALVYFFYKKGKLPGRSGKQGITLPPTRKSEFVEVKSCLKP 60
QY 619 EEMGLQGSSGDKRAPDQGEKYIDLRH 646
DB 61 EEMALLQGSNGDKRAPDQGEKYIDLRH 88

RESULT 13
019129 PRELIMINARY; PRT; 86 AA.
AC 019129;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE MELANOMA CELL ADHESION MOLECULE (FRAGMENT).
GN MCAM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A.
RA TISSUE-SKELETAL MUSCLE;
RC Moore S.S., Byrne K.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U89328; AAB66511.1; -.
FT NON-TER
SQ SEQUENCE 86 AA; 9486 MW; F109345FDB2B3555 CRC64;

Query Match 10.8%; Score 362; DB 6; Length 86;
Best Local Similarity 93.3%; Pred. No. 3,5e-21;
Matches 70; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 572 LAVGAVLYFLYKKGKLPGRSGKEITLPPSRKSELVEVKSCLKPREMGLQGSSGDK 631
DB 12 LGLLAVAVLYFFYKKGKLPGRSGKEITLPPSRSEFVEVKSCLKPREMGLQGSSGDK 71
QY 632 RAPDQGEKYIDLRH 646
DB 72 RAPDQGEKYIDLRH 86

RESULT 14
09M208 PRELIMINARY; PRT; 355 AA.
AC 09M208;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE LUTHERAN GLYCOPROTEIN (FRAGMENT).
GN LU.
OS Bos taurus (Bovine).

```


CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Boado R.O., Li J.Y., Pardridge W.M.;
 RT "Selective Lutheran glycoprotein gene expression at the blood-brain
 barrier in normal brain and in human brain tumors."
 RL J. Cereb. Blood Flow Metab. 0:0-0(2000).
 FT EMBL: AF270512; AAF81749.1; -
 SO NON_TER 1
 SEQUENCE 355 AA; 38047 MW; BEFB9772F6427EC6 CRC64;
 Query Match 10.7%; Score 361; DB 6; Length 355;
 Best Local Similarity 28.6%; Pred. No. 3.3e-20;
 Matches 114; Conservative 63; Mismatches 155; Indels 66; Gaps 15;
 QY 260 GMLKSGDVEIRCLADGNPPPHFSI---SKNPSTREAEETNDNGVLLEPARKHS 315
 DB 6 GMLKSGDVEIRCLADGNPPPHFSI---SKNPSTREAEETNDNGVLLEPARKHS 315
 QY 316 GRYCQGLDPTMISL-LSEPQELVNVSDVRVSPAPRQE---GSSLTTCCEASSQ 371
 DB 60 GRYCQGLDPTMISL-LSEPQELVNVSDVRVSPAPRQE---GSSLTTCCEASSQ 371
 QY 372 DLRFQWIREFTGVLENGPVLOLHDLKREAGGRCVAPSPITGLNRTQVNVATFGPP 431
 DB 118 TPLTYWTKDAS--PMGEDPTLUSHVTFDSAGYTCVAMPRIPLSLRTRSRFLVQGRP 117
 QY 432 WMAEKERK---VWVENNVNLSCASGHPPTISMN-VNCTASPDODDOPR-VLSTL 484
 DB 176 ELKAKETQPAEAGSWTEGDEV-FLICVARGYPRPKLWLSQSGSPTRPAGCGGWSSSL 234
 QY 485 NVLTPBELLETGECTASNDLGKNTSLLELVNLTLPDSNTTGLSTFASPTHRAN 544
 DB 235 TLKTSALSDQGVSCASNPLG-----WTHHVFHFQVAPQT--- 271
 QY 545 STERERKLPEPESRGVYIVVICILVAVLAVLFLYKKGKLPCKRSKQKBITLPPSR 604
 DB 272 -----SOAGVAVMVAISVALLLVAVVFCMRKGRPGCCQWG--EKGSPPPG 318
 QY 605 KSELVEVSKDKLPBENGLLQSS--GDKRAPGDQGER 640
 DB 319 EPKL--SHSGSQREQGTGLMGASAGAKHGSGGFQDE 354
 RESULT 15
 076518 PRELIMINARY; PRT; 5198 AA.
 AC 076518; Q10036;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE HEMICENTIN PRECURSOR.
 GN HIM-4 OR F15G9.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-BRISTOL N2;
 RC Vogel B.E., Hedgecock E.M.;
 RT "Hemicentlin is Required for Hemidesmosome Mediated Cell Adhesion and
 RT Germ-Line Chromosome Segregation in C. elegans."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN-BRISTOL N2;
 RC Sulston J., Kershaw J.;
 RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.

CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, ISOFORM F15G9.4A AND ISOFORM
 CC F15G9.4B (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
 DR EMBL: AF074901; AAC26792.1; -
 DR EMBL: 247068; CAAB7336.1; -
 DR EMBL: 247070; CAAB7336.1; JOINED.
 DR EMBL: 247070; CAAB7345.1; -
 DR EMBL: 247068; CAAB7345.1; JOINED.
 DR EMBL: 247068; CAAB7335.1; -
 DR EMBL: 247070; CAAB7335.1; JOINED.
 DR EMBL: 247070; CAAB7344.1; -
 DR EMBL: 247068; CAAB7344.1; JOINED.
 DR HSSP: P35555; IEMN.
 DR WORMREP: F15G9.4A; CE18595.
 DR WORMREP: F15G9.4B; CE18596.
 DR INTERPRO: IPR000152; -
 DR INTERPRO: IPR000515; -
 DR INTERPRO: IPR000561; -
 DR INTERPRO: IPR001881; -
 DR INTERPRO: IPR003006; -
 DR PFAM: PF00047; 19; 47.
 DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_1.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 DR PROSITE: PS01187; EGF_CA_2.
 KW Signal; Glycoprotein; EGF-like domain; Alternative splicing.
 FT SIGNAL 1 24
 FT CHAIN 25 5198
 FT VARSPIC 1462 5198
 FT SEQUENCE 5198 AA; 570816 MW; DA8511FE2858D37B CRC64;
 Query Match 8.4%; Score 281; DB 5; Length 5198;
 Best Local Similarity 22.1%; Pred. No. 3e-12;
 Matches 122; Conservative 91; Mismatches 220; Indels 118; Gaps 27;

QY 27 EAEQAPAELEVEVEGSTAL-----LKGLSQSGCNLSHVDFSVHKKRKLIRVRQG- 79
 DB 1357 DVQEPFILLPSTQTNNTAVAGGRVLCYKVENSP--ASVWF-----RGT 1401
 QY 80 -GGSEPEGEYRQLSLDQKATLALQVTPDERLFLCGKRRPSQ--EYRIQLRVYAPE 137
 DB 1402 AIGFTPTKG-----YVSSDGLVLIQSASVEDATITCKASNPAGKANQVNVIASPD 1455
 QY 138 --EPMQVNPGLIPVNSKPEPEVATCVGRNGYPIQVYIWKNGRL-----KEKN 186
 DB 1456 IDPQVYVQE---STKESHPESLYCPRFSN--PLQIISWYINDKPLIDDKTSWTSDDKR 1510
 QY 187 RVHIQSQTVESSGLYTTLQSLIKALQVLEKEDKDAQCYCELNRLPSGNHMKSR--EYTVPV 245
 DB 1511 KIHVFRKAK-IIDSGYKKVA-----RNAAGSGSKSFQVEYIVPL 1548
 QY 246 EYPTPKVWLEVPVGMKEDGVEIRCLADGNPPPHFSI---SKNPSTREAEETND 301
 DB 1549 NLDESKYKKYK---FAKEGEVTLGCPVSGFPVQIMWVDGIVVEBKKYKATLSND 1604
 QY 302 NCVLLEPARKHSGRYCQG-----LDLPTMISLSEPQELVNVSDVRVSPAPR 355
 DB 1605 GILTLFDSVSKQEBNHYCAVQSKNIIIDVIELSVLAVP---IVGEDDNLVYF----- 1655
 QY 356 QEGSSLTLTC--EAESSDLEFQWL-----REFTGVLENGPVLOLHDLKREAGGGR 406
 DB 1656 -LGDISLSCDLQTFSDSKITFWWSINGSESDRDQVQIPDGDHRLVITTAKEPNNQKYM 1714
 QY 407 CVASVPSTIPGLNRQVLNV---AIFGPPWMAFKERKWKVKNVNLSCASGHPPTIS 463
 DB 1715 CRVT--NSAGKAERLLTLDVLEPPVFEV--VEFANOKLIGNNPII-LQCVGTGNPKPTVI 1770
 QY 464 WNVNCTAGEODQDQPVLTSLNVVLTPELLEPTG---VECTASNDLGKNTSLILELVNLT 519
 DB 1771 WKIDQNDVDRKSWLDEESILSLRI-----EKL-TGSAQISCTAENKAGTASRDFPIONIAA 1825
 QY 520 TPLPDSNTTT 530

Mon Mar 26 09:22:40 2001

us-09-653-961-2.rspt

Page 10

Db 1826 PTFKNEGDET 1836

Search completed: March 23, 2001, 12:32:35
Job time: 113 sec